Cross-Modal Dynamic Attention Neural Architecture

Subjects: Computer Science, Artificial Intelligence

Contributor: Konstantinos Demertzis , Konstantinos Rantos , Lykourgos Magafas , Lazaros Iliadis

Detecting anomalies in data streams from smart communication environments is a challenging problem that can benefit from novel learning techniques. The Attention Mechanism is a very promising architecture for addressing this problem. It allows the model to focus on specific parts of the input data when processing it, improving its ability to understand the meaning of specific parts in context and make more accurate predictions.

cross-modal learning tasks | dynamic attention mechanism | neural architecture

1. Introduction

Detecting anomalies in data streams [1] from smart communication environments is a critical problem that has significant implications for various applications, including cyber security [2], monitoring cyber-physical systems [3], and controlling the industrial ecosystem [4]. The vast amount of data generated in these environments makes it difficult to detect abnormal behavior in real-time, which can lead to significant damages and security breaches [5]. Anomaly detection in these data streams is challenging due to the volume and complexity of the data and the need for real-time detection to prevent potential damages or security breaches [6][7]. Traditional methods for anomaly detection in data streams rely on statistical techniques or rule-based systems, which may not be effective in identifying subtle or unknown anomalies [8]. Machine learning approaches, particularly deep learning methods, have shown promise in addressing this challenge by enabling automated and accurate detection of anomalies in complex data streams [9].

One of the key advantages of deep learning methods for anomaly detection is the ability to learn relevant features from the input data without relying on pre-defined rules or assumptions. Attention mechanisms, in particular, have emerged as a powerful tool for capturing relevant input data features and improving neural network performance in various applications [2]. Recent research has focused on developing novel deep-learning architectures that effectively leverage attention mechanisms to detect anomalies in data streams from smart communication environments. These architectures often use simple attention mechanisms that can adapt to changes in the input data over time and can be trained end-to-end using data streams to capture the complex interactions between sophisticated processes [10][11].

Simple attention involves computing a fixed set of attention weights for the input data learned during training based on the task-specific objective function. The network then uses these fixed attention weights to weigh the input features in subsequent neural network layers. These simple attention mechanisms have become a powerful tool for capturing relevant input data features and improving neural network performance in various applications [12].

On the other hand, dynamic attention allows the network to adjust the attention weights at each time step to give more or less importance to different parts of the input sequence depending on their relevance to the task. Dynamic attention mechanisms can be useful in applications where the types and frequencies of anomalies may change over time, allowing the model to adapt to changes in the input data [13].

Both simple and dynamic attention mechanisms have strengths and weaknesses depending on the specific application and data. Simple attention is more straightforward and can be effective in many cases. In contrast, dynamic attention can improve the model's ability to adapt to changes in the input data over time. The appropriate attention mechanism type depends on the input data's nature and task [14][15].

2. Cross-Modal Dynamic Attention Neural Architecture

Anomaly detection in data streams has been an active research area due to the increasing volume and complexity of data generated by IoT devices and smart environments [2]. Traditional anomaly detection methods, such as statistical techniques [5], clustering [16], and classification [8], have been applied to data streams [6], with varying degrees of success. However, they often struggle to adapt to the dynamic nature of data streams, which may have changing distributions and evolving patterns [5]. For example, during a timed event, the traffic pattern can change dramatically, potentially causing statistical methods that rely on historical data to label the surge in traffic as an anomaly due to the shift in statistical properties like mean and variance [17]. In addition, the traditional clustering methods might not recognize the sudden appearance of a new cluster as an anomaly, leading to delayed detection, or traditional classifiers might struggle to identify novel patterns that were not present in the training data [6]. In summary, traditional anomaly detection methods have limitations that become more pronounced in dynamic data streams with changing distributions and evolving patterns. The technical challenges of concept drift [17], high-dimensional data [7], computational efficiency [18], and feature engineering [19] contribute to their struggles in adapting to these scenarios. This has prompted the exploration of more advanced techniques, including deep learning-based approaches, which have shown better adaptability and scalability in handling the dynamic nature of data streams.

Recently, deep learning-based techniques [2] have been proposed for data stream anomaly detection, including autoencoders [20], recurrent neural networks (RNNs) [21], and convolutional neural networks (CNNs) [22]. These methods have demonstrated better adaptability and scalability compared to traditional methods, but they still face challenges in dealing with heterogeneous data types and efficiently focusing on relevant features. Specifically, deep learning techniques face significant challenges in dealing with heterogeneous data types and efficiently focusing on relevant features [2]. These challenges include handling diverse data types, ensuring feature relevance and selection, addressing data imbalance, and interpreting deep models [23]. Heterogeneous data types, such as numerical, categorical, text, image, and time series data, can be challenging to integrate and process effectively [7][24]. Researchers are exploring techniques to handle multiple data types [25], such as specialized network architectures [26] or converting different data types into a common feature space [27]. Feature engineering and selection techniques aim to identify the most informative features, while data imbalance can lead to models favoring the majority class and performing poorly in anomaly detection [28]. Interpretable models are crucial to understanding the underlying patterns learned by deep learning models, such as in manufacturing processes where engineers need to know which factors contributed to anomaly detection [29]. Researchers are developing techniques to explain deep model decisions, such as attention mechanisms, feature attribution methods, and gradient-based visualizations, to provide insights into which features were influential in making anomaly predictions [30].

Cross-modal learning [31] refers to the process of learning shared representations from multiple data modalities, such as images, text, and audio. It has shown great potential in various applications, including multimedia retrieval [32], recommendation systems [33], and multimodal sentiment analysis [25]. Several methods have been proposed for cross-modal learning, including deep neural networks [34], matrix factorization [35], and probabilistic graphical models [36]. Recently, cross-modal learning has been integrated with attention mechanisms to improve the interpretability and performance of the learned representations [37][38][39]. However, the application of cross-modal learning to anomaly detection in data streams from smart communication environments is still relatively unexplored. This approach offers several benefits, but also presents challenges, such as developing effective fusion strategies, addressing domain-specific issues, dealing with varying data modalities, and managing computational complexity [36]. Additionally, data privacy and ethics are critical concerns in smart communication environments, and researchers must address these concerns when designing cross-modal anomaly detection systems [25].

Attention mechanisms have been introduced in neural networks to help the model focus on the most relevant parts of the input data for a specific task [12]. The concept of attention was initially proposed in the context of Natural Language Processing (NLP) [15] and has since been extended to various domains, such as computer vision [14] and speech recognition [40]. Different types of attention mechanisms have been proposed, including self-attention [41], local attention [42], and global attention [43]. Attention mechanisms have also been combined with other neural network architectures, such as RNNs [44], CNNs [45], and Transformer models [46], to improve their performance and interpretability. The application of attention mechanisms in anomaly detection has shown promising results, particularly in terms of handling large-scale and high-dimensional data [27]. However, incorporating dynamic attention mechanisms into cross-modal learning for anomaly detection in data streams remains a challenge. Specifically, incorporating dynamic attention mechanisms into cross-modal learning for anomaly detection in data streams requires a careful balance between adaptability, efficiency, interpretability, and performance [37]. Researchers need to devise novel approaches that address these challenges and tailor dynamic attention mechanisms to the specific requirements of dynamic data streams and multi-modal data fusion [14]. Despite the challenges, successfully implementing dynamic attention can significantly enhance the accuracy and robustness of anomaly detection systems in complex and rapidly evolving environments [12].

In summary, research gaps from the literature review in anomaly detection in dynamic environments include adapting traditional methods to handle changing distributions and patterns, integrating heterogeneous data types, improving the interpretability of deep models, exploring cross-modal anomaly detection, incorporating dynamic attention mechanisms, and addressing privacy and ethics concerns. These areas highlight opportunities for innovation and exploration in anomaly detection in smart communication environments, particularly in integrating heterogeneous data types, enhancing interpretability, and effectively utilizing dynamic attention mechanisms and cross-modal learning techniques.

By addressing these gaps, the proposed approach proposes a more effective anomaly detection method that can handle diverse data types, improve interpretability, and maintain privacy and ethics in cross-modal anomaly detection systems. Specifically, this research presents a novel CM-DANA for detecting anomalies in data streams generated from smart communication environments. The proposed architecture leverages the advantages of cross-modal learning and dynamic attention mechanisms to effectively analyze heterogeneous data streams from different cyber modalities and identify anomalous patterns in real-time. Recent advancements inspire this approach in cross-modal learning and attention mechanisms in neural networks. Cross-modal learning has shown its potential in various applications where data comes from multiple sources or modalities, while attention mechanisms have been successful in helping models focus on relevant parts of input data for specific tasks. By combining these two concepts, the proposed approach not only improves the overall performance of anomaly detection but also enhances the interpretability and adaptability of the model in handling diverse and evolving data patterns.

The proposed method addresses research gaps in anomaly detection in dynamic data streams from smart communication environments by enhancing traditional methods, integrating heterogeneous data types, enhancing interpretable deep models, incorporating cross-modal learning, and incorporating dynamic attention mechanisms. These contributions can help develop more accurate, adaptive, and interpretable anomaly detection systems that can effectively operate in complex and rapidly evolving scenarios. By incorporating concepts from both the dynamic attention and anomaly detection domain, the proposed CM-DANA technique ensures that data from different modalities are integrated in an accurate way. By focusing on these contributions, the proposed approach makes significant strides in advancing the field of anomaly detection in dynamic data streams from smart communication environments.

References

1. Golab, L.; Ozsu, M.T.; Data Stream Management. Morgan & Claypool. 2010. Available online: https://books.google.gr/books/about/Data_Stream_Management.html?id=IMyogd_LF1cC&redir_esc=y (accessed on 22 July 2020).
2. Dawoud, A.; Shahristani, S.; Raun, C. Deep Learning for Network Anomalies Detection. In Proceedings of the 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, Australia, 3–7 December 2018; pp. 149–153.
3. Jara, A.J.; Genoud, D.; Bocchi, Y. Big Data for Cyber Physical Systems: An Analysis of Challenges, Solutions and Opportunities. In Proceedings of the Eighth International Conference

on Innovative Mobile and Internet Services in Ubiquitous Computing, Birmingham, UK, 2–4 July 2014; pp. 376–380.

4. Ali, R.F.; Muneer, A.; Dominic, P.D.D.; Ghaleb, E.A.A.; Al-Ashmori, A. Survey on Cyber Security for Industrial Control Systems. In Proceedings of the 2021 International Conference on Data Analytics for Business and Industry (ICDABI), Online, 25–26 October 2021; pp. 630–634.
5. Vafaie, B.; Shamsi, M.; Javan, M.S.; El-Khatib, K. A New Statistical Method for Anomaly Detection in Distributed Systems. In Proceedings of the 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada, 30 August–2 September 2020; pp. 1–4.
6. Jirsik, T. Stream4Flow: Real-time IP flow host monitoring using Apache Spark. In Proceedings of the NOMS 2018—2018 IEEE/IFIP Network Operations and Management Symposium, Taipei, Taiwan, 23–27 April 2018; pp. 1–2.
7. Benjelloun, F.-Z.; Lahcen, A.A.; Belfkih, S. An overview of big data opportunities, applications and tools. In Proceedings of the 2015 Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 25–26 March 2015; pp. 1–6.
8. Guo, S.; Liu, Y.; Su, Y. Comparison of Classification-based Methods for Network Traffic Anomaly Detection. In Proceedings of the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 18–20 June 2021; pp. 360–364.
9. Dai, J.J.; Wang, Y.; Qiu, X.; Ding, D.; Zhang, Y.; Wang, Y.; Jia, X.; Zhang, C.L.; Wan, Y.; Li, Z.; et al. BigDL: A Distributed Deep Learning Framework for Big Data. In Proceedings of the ACM Symposium on Cloud Computing, Santa Cruz, CA, USA, 20–23 November 2019.
10. Gallicchio, C.; Micheli, A. Deep Echo State Network (DeepESN): A Brief Survey. arXiv 2019, arXiv:1712.04323.
11. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. arXiv 2018, arXiv:1808.01974.
12. He, W.; Wu, Y.; Li, X. Attention Mechanism for Neural Machine Translation: A survey. In Proceedings of the 2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Xi’an, China, 15–17 October 2021; pp. 1485–1489.
13. Tao, A.; Sapra, K.; Catanzaro, B. Hierarchical Multi-Scale Attention for Semantic Segmentation. arXiv 2020, arXiv:2005.10821.
14. Sun, J.; Jiang, J.; Liu, Y. An Introductory Survey on Attention Mechanisms in Computer Vision Problems. In Proceedings of the 2020 6th International Conference on Big Data and Information Analytics (BigDIA), Shenzhen, China, 4–6 December 2020; pp. 295–300.

15. Zhang, N.; Kim, J. A Survey on Attention mechanism in NLP. In Proceedings of the 2023 International Conference on Electronics, Information, and Communication (ICEIC), Singapore, 5–8 February 2023; pp. 1–4.
16. Deng, D. Research on Anomaly Detection Method Based on DBSCAN Clustering Algorithm. In Proceedings of the 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT), Shenyang, China, 13–15 November 2020; pp. 439–442.
17. Lu, J.; Liu, A.; Dong, F.; Gu, F.; Gama, J.; Zhang, G. Learning under Concept Drift: A Review. IEEE Trans. Knowl. Data Eng. 2018, 31, 2346–2363.
18. Cao, K.; Liu, Y.; Meng, G.; Sun, Q. An Overview on Edge Computing Research. IEEE Access 2020, 8, 85714–85728.
19. Wang, J.; Chen, J.; Lin, J.; Sigal, L.; de Silva, C.W. Discriminative feature alignment: Improving transferability of unsupervised domain adaptation by Gaussian-guided latent alignment. Pattern Recognit. 2021, 116, 107943.
20. Qin, K.; Zhou, Y.; Tian, B.; Wang, R. AttentionAE: Autoencoder for Anomaly Detection in Attributed Networks. In Proceedings of the 2021 International Conference on Networking and Network Applications (NaNA), Lijiang City, China, 29 October–1 November 2021; pp. 480–484.
21. Sokolov, A.N.; Alabugin, S.K.; Pyatnitsky, I.A. Traffic Modeling by Recurrent Neural Networks for Intrusion Detection in Industrial Control Systems. In Proceedings of the 2019 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), Sochi, Russia, 25–29 March 2019; pp. 1–5.
22. Liu, S.; Jiang, H.; Li, S.; Yang, Y.; Shen, L. A Feature Compression Technique for Anomaly Detection Using Convolutional Neural Networks. In Proceedings of the 2020 IEEE 14th International Conference on Anti-Counterfeiting, Security, and Identification (ASID), Xiamen, China, 30 October–1 November 2020; pp. 39–42.
23. Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Comput. Sci. 2021, 2, 420.
24. Tsimenidis, S.; Lagkas, T.; Rantos, K. Deep Learning in IoT Intrusion Detection. J. Netw. Syst. Manag. 2021, 30, 8.
25. Peng, C.; Zhang, C.; Xue, X.; Gao, J.; Liang, H.; Niu, Z. Cross-modal complementary network with hierarchical fusion for multimodal sentiment classification. Tsinghua Sci. Technol. 2022, 27, 664–679.
26. Sanla, A.; Numnonda, T. A Comparative Performance of Real-time Big Data Analytic Architectures. In Proceedings of the 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 12–14 July 2019; pp. 1–5.

27. Liu, F.; Zhou, X.; Cao, J.; Wang, Z.; Wang, T.; Wang, H.; Zhang, Y. Anomaly Detection in Quasi-Periodic Time Series Based on Automatic Data Segmentation and Attentional LSTM-CNN. IEEE Trans. Knowl. Data Eng. 2022, 34, 2626–2640.
28. Sani, Y.; Mohamedou, A.; Ali, K.; Farjamfar, A.; Azman, M.; Shamsuddin, S. An overview of neural networks use in anomaly Intrusion Detection Systems. In Proceedings of the 2009 IEEE Student Conference on Research and Development (SCOReD), Seri Kembangan, Malaysia, 16–18 November 2009; pp. 89–92.
29. Embarak, O. Decoding the Black Box: A Comprehensive Review of Explainable Artificial Intelligence. In Proceedings of the 2023 9th International Conference on Information Technology Trends (ITT), Dubai, United Arab Emirates, 24–25 May 2023; pp. 108–113.
30. Sasaki, H.; Hidaka, Y.; Igarashi, H. Explainable Deep Neural Network for Design of Electric Motors. IEEE Trans. Magn. 2021, 57, 1–4.
31. Xu, X.; Lin, K.; Gao, L.; Lu, H.; Shen, H.T.; Li, X. Learning Cross-Modal Common Representations by Private–Shared Subspaces Separation. IEEE Trans. Cybern. 2022, 52, 3261–3275.
32. Hua, Y.; Du, J. Deep Semantic Correlation with Adversarial Learning for Cross-Modal Retrieval. In Proceedings of the 2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 12–14 July 2019; pp. 256–259.
33. Tie, Y.; Li, X.; Zhang, T.; Jin, C.; Zhao, X.; Tie, J. Deep learning based audio and video cross-modal recommendation. In Proceedings of the 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Prague, Czech Republic, 9–12 October 2022; pp. 2366–2371.
34. Ma, M.; Liu, W.; Feng, W. Deep-Learning-based Cross-Modal Luxury Microblogs Retrieval. In Proceedings of the 2021 International Conference on Asian Language Processing (IALP), Yantai, China, 23–25 October 2021; pp. 90–94.
35. Liu, X.; Hu, Z.; Ling, H.; Cheung, Y.-M. MTFH: A Matrix Tri-Factorization Hashing Framework for Efficient Cross-Modal Retrieval. IEEE Trans. Pattern Anal. Mach. Intell. 2021, 43, 964–981.
36. Chun, S.; Oh, S.J.; de Rezende, R.S.; Kalantidis, Y.; Larlus, D. Probabilistic Embeddings for Cross-Modal Retrieval. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 8411–8420.
37. Wang, X.; Liang, M.; Cao, X.; Du, J. Dual-pathway Attention based Supervised Adversarial Hashing for Cross-modal Retrieval. In Proceedings of the 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju-si, Republic of Korea, 17–20 January 2021; pp. 168–171.
38. Fang, Z.; Li, L.; Xie, Z.; Yuan, J. Cross-Modal Attention Networks with Modality Disentanglement for Scene-Text VQA. In Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 18–22 July 2022; pp. 1–6.

39. Guan, W.; Wu, Z.; Ping, W. Question-oriented cross-modal co-attention networks for visual question answering. In Proceedings of the 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 14–16 January 2022; pp. 401–407.

40. Zhang, S.; Loweimi, E.; Bell, P.; Renals, S. Windowed Attention Mechanisms for Speech Recognition. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 7100–7104.

41. Kim, M.; Kim, T.; Kim, D. Spatio-Temporal Slowfast Self-Attention Network for Action Recognition. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 2206–2210.

42. Yan, H.; Zhang, E.; Wang, J.; Leng, C.; Liang, H.; Peng, J. Coarse-Refined Local Attention Network for Hyperspectral Image Classification. In Proceedings of the 2022 International Conference on Image Processing and Media Computing (ICIPMC), Xi'an, China, 27–29 May 2022; pp. 102–107.

43. Deng, S.; Dong, Q. GA-NET: Global Attention Network for Point Cloud Semantic Segmentation. IEEE Signal Process. Lett. 2021, 28, 1300–1304.

44. Shu, X.; Zhang, L.; Qi, G.-J.; Liu, W.; Tang, J. Spatiotemporal Co-Attention Recurrent Neural Networks for Human-Skeleton Motion Prediction. IEEE Trans. Pattern Anal. Mach. Intell. 2022, 44, 3300–3315.

45. Zhang, Z.; Jiang, T.; Liu, C.; Ji, Y. Coupling Attention and Convolution for Heuristic Network in Visual Dialog. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; pp. 2896–2900.

46. Jiang, Y.; Wang, J.; Huang, T. Prediction of Typhoon Intensity Based on Gated Attention Transformer. In Proceedings of the 2022 International Conference on High Performance Big Data and Intelligent Systems (HDIS), Tianjin, China, 10–11 December 2022; pp. 141–146.