

Research on the Application of AI Technologies in Computer Network Data Analysis

Ho Tsz Lok^{1,*}

¹Department of Computer Science, Hong Kong Baptist University, Hong Kong, China

*Corresponding author: timtszlok@gmail.com

Abstract. In recent years, with the development of the Internet of Things and smart devices, the Future Internet project has driven innovation in network architecture by combining different models and resources. To manage complexity, creating abstract models has become a critical step in the field of computer networks. This study explores the application of AI techniques in network data analysis, demonstrating the possibilities of AI for real-time monitoring, anomaly detection, and the future development of AI. Despite the promising applications of AI technology, challenges such as privacy protection and dealing with cyber threats still need to be overcome. Future research should develop safer and more flexible AI algorithms to adapt to complex network environments.

Keywords: AI, Data Analysis, Model, Data.

1. Introduction

This trend is particularly evident in the various endeavors of recent years that are applicable to describing not only the network itself but also the devices connected to it. With the rapid development of the Internet of Things (IoT) and smart devices, Future Internet projects are driving innovation in network architecture by combining different models and resources to provide users with a complete virtual infrastructure[1]. These projects are not only focused on connectivity at the physical level but also on building a highly integrated digital ecosystem that enables users to access the services and information they need, wherever and whenever they need it. From now on, AI will be more widely used in data analytics.

1.1. Significance of the Study

Creating abstract models in order to manage complexity is one of the important steps taken in the field of computer networks. The increasing number of devices with networking capabilities, coupled with the growing popularity of the Internet, has made computer networks a key area of modeling. The presence of many connected devices increases the complexity of the network and must be controlled to ensure its proper functioning. Therefore, it is of great practical importance to investigate how to use advanced technological tools, especially AI, to simplify network management tasks, improve network performance, and enhance network security and protection.

1.2. Objectives

A variety of computer network models for different purposes have been proposed over the years. This research aims to explore how AI techniques can be applied to address the current challenges facing networks, particularly in terms of efficient analysis of network data. Issues such as cybersecurity, which have arisen as the number and size of networks have grown, suggest that current networking technologies, although numerous, still need to address critical issues such as cybersecurity fully. As networks become sophisticated and serve as tools for everyday social and business interactions, old strategies for network development are no longer applicable.[2] Therefore, a new network model is needed to address current pressing issues, especially those related to cybersecurity. By introducing AI technology, this study hopes to provide new ideas and solutions for network data analysis and help build a more secure and efficient network environment.

2. Artificial Intelligence (AI) and its Applications

2.1. Overview of AI Technologies

What is AI? AI is a branch of the field of computer science dedicated to the development of machine systems capable of demonstrating human-like intelligent behavior[3]. These intelligent programs continuously improve their ability to mimic human behavior by collecting and analyzing data from various interactions, thus enabling functions such as learning, planning, knowledge representation, perception, and problem-solving. Their applications are extremely wide-ranging and can be used, for example, in web development to create automated chatbots for customer service, to provide product recommendations based on user habits, to implement speech recognition functions, and even to build websites fully automatically. Overall, the goal of AI is to optimize the workflow of existing systems and increase their efficiency through automation.

Artificial intelligence systems use algorithms to process huge amounts of data in order to recognize patterns in it and predict the future behavior of end users accordingly, thus helping the system to understand better and simulate human intelligence. For example, in chatbots that utilize AI technology, the system is able to anticipate and provide information in advance in response to the next question a user may ask, just like a real chat agent[3].

2.2. Applications of AI in Other Fields

AI is used in a wide range of fields nowadays. In automated systems, AI applications are often developed and integrated as distributed systems involving multiple experts, each using their domain-specific modeling language and tools to model system elements. The lack of a cross-disciplinary graphical modeling language makes it difficult to develop a unified understanding, which in turn increases the cost of development, integration, and maintenance. To address this problem, researchers have proposed a graphical modeling language (GML) that models AI applications in a consistent and easy-to-understand system-level manner.

This new GML defines notations, rules of combination, and their semantics that allow modeling system components, functions, and their relationships, thus enabling the graphical representation of elements of automated systems and AI applications, even under different AI system architectures, such as cloud or edge architectures. Furthermore, with GML, the interfaces between components and the distribution of the software on the hardware become transparent to all participants and experts. As a result, this graphical representation provides a communication framework for interdisciplinary collaboration, facilitating communication while documenting the solution and making it easy to understand. GML can be used in parallel with common process models (e.g., CRISP-DM), thus minimizing the additional workload associated with the use of GML[4].

The goal of GML is to simplify and standardize the documentation of AI applications in automated systems, improving their subsequent integration and long-term maintenance throughout their lifecycle. Currently, GML still needs to be able to model or represent temporal dependencies between functions; e.g., data preprocessing needs to be done before reasoning. It is also yet to be possible to assign attributes to individual system elements, which would allow documenting and displaying important system properties, such as communication speeds used, communication protocols, or data-driven algorithms.

Over the past few years, with the successful development of various AI approaches, the field of particle radiotherapy has been able to take advantage of AI technology to improve the accuracy and efficiency of its clinical workflow. Although there have been many reviews of the use of AI in radiotherapy, this is the first comprehensive review of the use of AI specifically for particle therapy. The use of AI technology in particle therapy can be divided into four main categories: particle therapy planning, adaptive particle therapy, range and dose validation, and other applications[5]. In particle therapy planning, AI techniques such as deep learning have been used to generate synthetic dual-energy CT images from conventional CT images and thus obtain more accurate relative stopping

power (SPR) maps for dose calculation. For adaptive particle therapy, AI techniques have helped enable adaptive treatment plan adjustment based on cone beam CT (CBCT) images. For range and dose validation, AI was also applied to verify the actual dose distribution during treatment via PET images. Although AI-based approaches show a lot of promise, there are still some challenges to be overcome to scale up these technologies in practical clinical applications, such as the need for larger validation studies and the resolution of some practical issues during clinical implementation.

With the development of 5G mobile communication networks, there are many more areas where the application of AI has paved the way, and these applications include innovative features such as holographic communication and high-precision manufacturing. 5G technology increases communication capacity, reduces transmission latency, and lowers error rates, all of which are needed for emerging applications. Examples include immersive video conferencing, augmented reality, virtual reality, self-driving vehicles, disaster response, healthcare, education, high-precision manufacturing, and many other areas[6]. To support these applications, future research directions should encourage interdisciplinary collaboration to address the trade-offs between security, privacy, performance, and cost.

3. Artificial Intelligence Implementation Process

3.1. Data Collection and Preprocessing

A network data model is a data representation that focuses on the connections and interactions between entities, providing a dynamic perspective for understanding the complex relationships within a system[7]. Unlike traditional hierarchical or relational data models that rely on tabular or tree structures, the network data model adopts a more flexible and highly interconnected approach to organizing data. The main characteristics of network data are reflected in several aspects: firstly, the huge amount of data generated and transmitted by the network every day puts high demands on the system's scalability and storage capacity; secondly, there are a variety of network data types covering text, image, video, audio and other formats, which increases the processing difficulty; in addition, the data generation and update speed is extremely fast, which requires the system to have a real-time or near-real-time processing capability; Furthermore, network data often contains a large amount of noise information, which needs to be cleaned before analysis; network data is mostly in unstructured or semi-structured form, such as posts on social media and web content, which requires a special method to extract valuable information from it.

Data preprocessing, on the other hand, is part of data preparation and involves various types of processing of raw data aimed at making it suitable for subsequent data processing procedures. This aspect has traditionally been seen as an important beginning of the data mining process. And in recent times, data preprocessing techniques have been applied not only for training machine learning and artificial intelligence models, but also for the inference process of these models. Through preprocessing, data is transformed into a form that is easier to handle in data mining, machine learning, and other data science tasks. These techniques are usually applied in the initial stages of the machine learning and AI development process to guarantee the accuracy of the results[8].

3.2. Selection of AI Models

Before choosing an AI model, one needs to know how to use AI for data analysis.

1. Data collection: Data is extracted from different data sources to feed AI algorithms.
2. Data cleaning: The cleaner the data, the more valuable the insights.
3. Data analysis: After training an AI model using clean, relevant data, you can start analyzing the data and generating actionable insights.
4. Data visualization: Build visual dashboards to support decision-making. Interactive charts and graphs allow you to explore data in depth and drill down to specific information to improve workflow.

5. Predictive analytics: Based on historical data, run predictive models to make accurate predictions.
6. Data-driven decision-making: With these powerful insights, make better decisions faster[9].

We now have a very large number of models or AI platforms that can do this. AI models:

1. Exponential families of random graph models (ERGMs)
2. Stochastic block models (SBMs)
3. Latent position models (LPMs)

AI platforms and tools:

1. Luzmo
2. Instachart
3. Tableau
4. Microsoft Power BI
5. KNIME
6. Databricks
7. AnswerRocket

3.3. Case Studies of Model Applications

Exponential families of random graph models (ERGMs): The model includes terms that control for structural effects such as edges, reciprocity, and out-degree terms in the network, as well as terms that adjust for gender-based edge effects and homogeneity. In addition, the geometrically-weighted shared partner (GWESP) term is included, which is intended to capture a stochastic tendency of network transmissibility and triangle formation based on the underlying parameter values and decay parameters. In an application to the Polish school network dataset, the highest selection probabilities were obtained for the model M4 containing the estimated attenuation parameters, which is in line with previous findings.

Stochastic block models (SBMs): The network data is modeled by considering a different number of blocks ($K = 1, 2, 3, 4$). These models are used to capture the community structure in the network. In the application to the Sampson monks network dataset, the models were evaluated by comparing different numbers of blocks, and finally, a model with three blocks was chosen.

Latent position models (LPMs): The model considers network density and reciprocity modeling terms, as well as different dimensions of potential space ($K = 1, 2, 3, 4$). Similarly, the latent location model was applied to the Sampson Monk Network dataset to assess the latent structure in the network data.

4. Advanced Applications in Maintenance and Security

4.1. Network Monitoring and Security

With the increasing demand for traffic management, Traffic Management Centers (TMCs) are under immense work pressure as they are required to monitor large volumes of real-time traffic through visual camera systems. Injecting a degree of automation can reduce the pressure on human operators and help make timely decisions to reduce the impact of accidents and traffic congestion.

There are automated traffic monitoring systems based on deep convolutional neural networks and standalone graphical user interfaces. The system uses a range of advanced deep-learning algorithms to automate different traffic surveillance needs. For example, massively labeled video surveillance data is used to train models to detect queues, track stationary vehicles, and count the number of vehicles. The system uses pixel-level segmentation methods to detect traffic queues and predict their

severity. Also, by combining real-time object detection algorithms and different tracking systems, the system is able to detect broken-down vehicles and perform vehicle counting automatically.

The system is able to maintain good performance in different environments and does not drastically affect the results due to blurred camera views, low light, rain, and snow. The system not only detects static vehicles but also tracks the movement patterns of different kinds of vehicles and identifies dangerous driving behaviors such as reckless lane changing. During the experimental phase, the system demonstrated its effectiveness and was shown to function effectively under a wide range of conditions[11].

AI technology can also be applied in remote power network condition monitoring. Given that single wire earth return (SWER) networks in rural areas suffer from problems such as poor voltage regulation and HIFs, which can cause severe forest fires such as the Black Saturday event of 2009, the situation can be improved by using AI techniques to improve the condition monitoring of the grid. They developed a distributed online monitoring platform that combines power quality monitoring, real-time HIF identification, and transient classification. Useful features are extracted from current and voltage signals and an AI-based classification technique is developed to identify faults and transient events. Experiments show that the method has high HIF detection accuracy (98.67%) and low detection delay (115.2 ms)[12].

There is also a load identification method for identifying remote consumer load types, which detects load types from startup transients and proposes an edge-computing architecture to minimize data transfer for efficient load identification. The method exhibits an average 98% recognition accuracy in real-world applications. By doing so, rural grids can be better managed and maintained, and risks arising from power failures can be reduced.

4.2. Fault Detection and Prediction

Traditional network intrusion detection systems (IDSs) typically require complete attack information when detecting an attack, which means that the attack has already been executed on the target system and may have caused damage. To prevent this from happening, use an end-to-end early intrusion detection system designed to prevent cyberattacks, avoid further damage to systems, and prevent unplanned downtime and disruption.

The system uses a deep neural network-based classifier for attack identification. Unlike most correlation methods that rely on a manual feature selection process, the network is trained in a supervised manner to extract relevant features directly from raw network traffic data. In addition, the article introduces a new metric, ‘earliness’ (early detectability), to evaluate how far ahead the proposed method is in detecting attacks. Empirical evaluation of the CICIDS 2017 dataset shows that the method performs well and achieves an overall balanced accuracy of 0.803. With this approach, the system is able to reliably identify attacks at an early stage of processing traffic, thus enabling network administrators to take faster action to block traffic and issue alert messages or deploy countermeasures[13].

5. Perspectives on Current Limitations and Future Prospects

5.1. Limitations

1. In statistical network analysis, while there are many different models for network data, it remains a challenge to make comparisons between models with different mathematical underpinnings in order to select the best model. Current methods may not be effective in assessing how good or bad a model is, especially if the underlying principles of the models vary widely. New non-parametric methods can be developed to utilize the spectral information of the graph Laplace operator to obtain a set of well-defined goodness-of-fit measures for network data models. Alternatively, the applicability of the new methods to various classes of network data models could be explored to improve their usefulness in practical applications[10].

2. Traditional traffic surveillance relies on human labor, which is not only inefficient but also makes it difficult to respond in a timely manner when faced with large amounts of data. Despite the advantages of video surveillance systems in event recognition, environmental factors such as blurred images can pose a challenge to most models. If this problem is to be solved, deep learning techniques can be used to automate the processing of real-time traffic videos and improve surveillance efficiency. Multi-channel deep feature learning algorithms are applied to detect abnormal traffic flows and ensure that they can still work effectively under different conditions[11].

3. Remote power networks face problems such as inefficient voltage regulation and HIFs from SWER networks, which can lead to catastrophes such as forest fires. Much of the existing research has focused primarily on improving algorithmic accuracy, ignoring feasibility issues posed by computational complexity and the need for high-frequency signal inputs in practical applications. AI monitoring algorithms that can operate in resource-constrained environments can be developed while improving monitoring accuracy. Validate the performance of algorithms in real-time environments, especially for deployment on low-power devices[12].

5.2. Future Developments

The use of deep learning models allows features to be automatically extracted from raw data, reducing the reliance on manual feature selection and thus improving the accuracy and adaptability of the system. The future direction of development lies in building smarter automated systems that can not only respond and adapt quickly to changing situations but also learn and optimize themselves through technologies such as deep learning to better serve a wide range of fields such as power networks, traffic management, and cyber security. In addition, with advances in algorithms and technology, we will continue to explore ways to reduce the system's demand for computing resources and improve its scalability and utility while ensuring accuracy in the future.

6. Conclusion

This study, through an in-depth exploration of recent developments in the field of computer networks, reveals that with the advancement of various smart devices and IoT technologies, the complexity of the network itself has been increasing. In the face of this complexity, it is recognized that the creation of effective abstract models is not only key to understanding network behavior but also fundamental to the management and maintenance of large-scale network infrastructures. In this context, AI technology has shown great potential in network data analysis due to its powerful data processing capabilities and pattern recognition.

Despite the promising application of AI technology in network data analysis, there are still some challenges and limitations. Future research should aim to develop safer and more flexible AI algorithms to adapt to the increasingly complex cyber environment. In addition, interdisciplinary cooperation will be key to driving the application of AI technology in the cyber domain, where complex problems in network management can be better understood and solved by combining knowledge from computer science, statistics, and social science.

References

- [1] Ghijssen, Mattijs, et al. "Trends in Computer Network Modeling Towards the Future Internet." ArXiv, 2014, /abs/1402.3951. Accessed 5 Jan. 2024.
- [2] Singh, Ram K., and T. Ramajujam. "New Model of Network- a Future Aspect of the Computer Networks." ArXiv, 2009, /abs/0912.3985. Accessed 5 Jan. 2024.
- [3] Ricart, Judit Ruiz. "What Is Artificial Intelligence (AI)? AI Definition." Wix, wix-encyclopedia, 26 Mar. 2024, www.wix.com/encyclopedia/definition/artificial-intelligence?utm_source=google&utm_medium=cpc&utm_campaign=21361708472%5E162921400203%5Esearch+-+dsa&experiment_id=%5E%5E701641900273%5E&gad_source=1&gclid=Cj0KCQjw0Oq2BhCCARIsAA5hubVPUUgCPTcsjg3qkGUthdvh1wPYDW89hPDj5uvwjfU5ZdonneF_eoaAIX5EALw_wcB.

- [4] Schieseck, Marvin, et al. "A Graphical Modeling Language for Artificial Intelligence Applications in Automation Systems." ArXiv, 2023, /abs/2306.11767. Accessed 22 Mar. 2024.
- [5] Wu, Chao, et al. "Applications of Artificial Intelligence in Particle Radiotherapy." ArXiv, 2021, /abs/2102.03061. Accessed 26 Mar. 2024.
- [6] Bertino, Elisa, and Sujata Banerjee. "Artificial Intelligence at the Edge." ArXiv, 2020, /abs/2012.05410. Accessed 26 Mar. 2024.
- [7] Yen, Leon. "What Is a Network Data Model? Examples, Pros and Cons." Datamation, 24 May 2024, www.datamation.com/big-data/what-is-a-network-data-model-examples-pros-and-cons/.
- [8] Lawton, George. "Data Preprocessing: Definition, Key Steps and Concepts." Data Management, TechTarget, 31 Jan. 2022, www.techtarget.com/searchdatamanagement/definition/data-preprocessing.
- [9] "Using AI for Data Analysis: The Ultimate Guide (2024): Luzmo." AI for Data Analysis: The Ultimate Guide for 2024, Luzmo, 17 July 2024, www.luzmo.com/blog/ai-data-analysis.
- [10] Hidalgo, Jairo I., and Jonathan R. Stewart. "Model Selection for Network Data Based on Spectral Information." ArXiv, 2023, /abs/2301.02871. Accessed 27 Mar. 2024.
- [11] Mandal, Vishal, et al. "Artificial Intelligence Enabled Traffic Monitoring System." ArXiv, 2020, /abs/2010.01217. Accessed 28 May. 2024.
- [12] Sirojan, Tharmakulasingam. "Artificial Intelligence Based Sensor Data Analytics Framework for Remote Electricity Network Condition Monitoring." ArXiv, 2021, /abs/2102.03356. Accessed 28 May. 2024.
- [13] Ahmad, Tanwir, et al. "Early Detection of Network Attacks Using Deep Learning." ArXiv, 2022, /abs/2201.11628. Accessed 1 June. 2024.