Anomaly Detection Algorithms for LAN Failure Prediction

Subjects: Automation & Control Systems | Telecommunications | Computer Science, Artificial Intelligence

Contributor: Leila Rzayeva, Ali Myrzatay, Gulnara Abitova, Assiya Sarinova, Korlan Kulniyazova, Bilal Saoud, Ibraheem Shayea

Predicting Local Area Network (LAN) equipment failure is of utmost importance to ensure the uninterrupted operation of modern communication networks. The utilization of machine learning algorithms, specifically decision trees and support vector machines (SVMs), for predicting LAN failures represents a groundbreaking approach in network management.

Keywords: machine learning methods ; random forest ; decision tree ; SVM ; LAN ; failure prediction

1. Introduction

Local area networks (LANs) serve as the backbone of modern organizations, facilitating seamless communication, data sharing, and resource access. Their reliability and uninterrupted operation are critical for sustaining daily business operations, and any unexpected LAN failure can lead to significant disruptions, financial losses, and diminished productivity. In this context, the ability to predict and preemptively address LAN failures is a paramount concern for network administrators and IT professionals. Studies conducted by Positive Technologies, as cited in references [1][2], indicate an 11% rise in incidents (encompassing attacks and network equipment failures) in the first half of 2019 compared to the same period in 2018 [3]. This suggests that the information systems of organizations from 2019 did not witness a reduction in their susceptibility to failures.

Traditional LAN management approaches have largely relied on reactive strategies, responding to issues as they occur. However, the increasing complexity of LAN infrastructures, coupled with the growth in data traffic and the emergence of diverse networked devices, necessitates a more proactive and intelligent approach. This paradigm shift has given rise to the field of LAN failure prediction, leveraging advanced technologies, such as machine learning, data analytics, and network monitoring to anticipate and prevent network disruptions before they occur.

The current market for monitoring systems is highly saturated, with various companies offering a range of solutions, including Network Olympus, Observium, Nagios, PRTG network monitor, Kismet, NeDi, and Zabbix [4][5][6][7][8]. While most of these systems are proprietary and require payment, there are also open-source solutions available with open-source code. It is noteworthy that most of the listed systems employ the simple network management protocol (SNMP) to acquire real-time information on the status of L2 network equipment, including the temperature of switch processors, the level of processing of requests between devices, incoming and outgoing traffic, power supply status, and other network device components. Access to this information enables the optimization of L2 network device performance through the application of machine learning algorithms for predictive analytics regarding possible future malfunctions [9][10].

The motivation behind LAN failure prediction is two-fold. Firstly, it seeks to mitigate the costs associated with downtime, which can range from the loss of revenue and productivity to damage to an organization’s reputation. Secondly, it aims to enhance network performance, ensuring that LANs operate at optimal levels, meeting the demands of modern data-driven businesses. Achieving these goals requires the development of sophisticated prediction models, the identification of critical failure indicators, and the integration of real-time monitoring and alerting systems.

The utilization of machine learning algorithms, specifically decision trees and support vector machines (SVMs), for predicting LAN failures represents a groundbreaking approach in network management. Unlike traditional rule-based systems, these algorithms offer a novel, data-driven approach that harnesses the power of pattern recognition and classification. Decision trees, for instance, can automatically identify crucial network features and decision points, creating a dynamic model that adapts to changing network conditions. SVM, on the other hand, excels at finding complex relationships within LAN data that might not be apparent to human operators.

The significance of employing these machine learning algorithms cannot be overstated. LAN failures can result in severe disruptions, impacting productivity and potentially causing financial losses for organizations. By leveraging decision trees and SVM, network administrators can proactively anticipate and prevent failures by identifying subtle warning signs and deviations from normal network behavior. This not only reduces downtime and maintenance costs but also enhances network reliability and user satisfaction. Furthermore, as LANs continue to evolve and expand in complexity, the ability of these algorithms to adapt and learn from new data sources becomes increasingly crucial, making them indispensable tools in the ever-demanding field of network management.

2. Anomaly Detection Algorithms for LAN Failure Prediction

Many researchers have explored the use of anomaly detection algorithms like isolation forests, k-nearest neighbor (k-NN), and one-class SVM for LAN failure prediction. These methods aim to identify unusual patterns or behaviors in network traffic that could signify an impending failure. In addition, deep neural networks, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been applied to LAN failure prediction tasks. These models excel at capturing intricate relationships in time-series network data, making them suitable for predicting subtle anomalies. Furthermore, some studies have combined multiple machine learning algorithms or models into ensembles to improve the accuracy and robustness of LAN failure predictions. Random forests and gradient boosting are commonly used ensemble techniques in this context. Moreover, with the advent of big data technologies and cloud computing, researchers have explored the scalability and efficiency of LAN failure prediction models. Cloud-based solutions allow for the real-time analysis of large volumes of network data. Finally, LAN failure prediction research is crucial for ensuring the stability and reliability of modern networks. The application of machine learning techniques, combined with advances in data collection and processing, continues to drive innovation in this field, with the ultimate goal of minimizing network downtime and enhancing overall performance.

In [11], the authors suggested using advanced forecasting algorithms, such as ARIMA and neural networks, for predicting future outcomes. The authors discussed the importance of model evaluation and selection, which involves comparing the performance of different forecasting models using metrics such as MAE and RMSE. They proposed using a remote monitoring and control system, which can enable real-time monitoring and control of the textile production process. They discussed using sensors and other monitoring devices to collect data, which can be transmitted to a central control system for analysis and decision-making.

In [12], the authors' methodology involves various technical terms and concepts, such as time series analysis, predictive modeling, and machine learning algorithms. Time series analysis is used to identify patterns and trends in historical data, which can then be used to develop predictive models. These models can be based on statistical methods, such as regression analysis, or machine learning algorithms, such as neural networks or decision trees. The authors also discussed the importance of data pre-processing and feature engineering in developing accurate forecasting models. This involves cleaning and transforming raw data into a format that is suitable for analysis, as well as selecting relevant features or variables that are likely to have a significant impact on the outcome. In addition, the study discussed the use of various performance metrics, such as mean-squared error (MSE) and mean absolute error (MAE), to evaluate the accuracy of forecasting models. These metrics are used to quantify the difference between predicted and actual values, and to assess the overall performance of the model.

The methodology proposed in [13] includes several steps for predicting the remaining useful life of hydraulic components. These steps include data pre-processing, feature selection, model training, and performance evaluation. Furthermore, data pre-processing involves cleaning and transforming the raw data to make it suitable for analysis. Several techniques have been used in this study such as data normalization, outlier removal, and missing value imputation for data pre-processing.

Feature selection is the process of selecting the most important features that contribute to the predictive model's accuracy. This step can be very important and improve the accuracy of the model's results. Some studies used several feature selection techniques, such as correlation-based feature selection and recursive feature elimination to select the most relevant features [12][13].

Model training [14] involves selecting an appropriate machine learning algorithm and training it on the pre-processed data. Decision trees, in essence, can solve both classification and regression issues. A decision tree is constructed by breaking it down into distinct subsets, known as leaf nodes. These branches represent different possibilities based on the dataset and offer a well-defined goal, while the root node signifies the optimal choice. Several algorithms have been used in [14], such as linear regression, decision trees, and random forests for model training. Performance evaluation involves

assessing the accuracy of the trained models. The authors used several metrics, such as mean absolute error, root mean-squared error, and the coefficient of determination to evaluate the model's performance. They also used a k-fold cross-validation technique to evaluate the model's generalizability.

Overall, the methodology proposed in [13][15][16][17] is well-explained and includes several technical terms related to machine learning and predictive maintenance. In addition, several techniques and algorithms have been used for data pre-processing, feature selection, model training, and performance evaluation. This methodology made the proposed system more reliable.

On the other hand, the Bayes modeling method is also beneficial for predictive failure analysis. As proposed in [18], this method employs possibilistic Bayes models. By using these models, a system is designed to aid the monitoring and control staff in detecting potential failures. It also aids in the planning of optimal programs for predictive maintenance.

Simultaneously, the logistic regression method is featured prominently in studies [18][19][20] for failure prediction. This approach is utilized to dissect the elements that contribute to communication failure and anticipate failures in the grid metering automation system.

References

1. Potapov, V.I.; Shafeeva, O.P.; Doroshenko, M.S.; Chervenchuk, I.V.; Gritsay, A.S. Numerically-analytical solution of problem gaming confrontation hardware-redundant dynamic system with the enemy operating in conditions of incomplete information about the behavior of participants in the game. J. Phys. Conf. Ser. 2018, 1050, 012062.
2. Storozhenko, N.R.; Goleva, A.I.; Tunkov, D.A.; Potapov, V.I. Modern problems of information systems and data networks: Choice of network equipment, monitoring and detecting deviations and faults. J. Phys. Conf. Ser. 2020, 1546, 012030.
3. Cybersecurity Threatscape 2019 (No Date) Ptsecurity.com. Available online: Https://www.ptsecurity.com/ww-en/analytics/cybersecurity-threatscape-2019/ (accessed on 19 August 2023).
4. Juliono, A.; Rosyani, P. Implementasi Sistem Monitoring Jaringan Internet Kantor PT. Permodalan Nasional Madani (Persero) Menggunakan Jessie Observium Dan Mikrotik (Simonjangkar). Kernel J. Ris. Inov. Bid. Inform. Pendidik. Inform. 2022, 3, 27–32.
5. Josephsen, D. Building a Monitoring Infrastructure with Nagios; Prentice Hall PTR: Indianapolis, IN, USA, 2007.
6. Zhou, J.; Huang, H.; Mattson, E.; Wang, H.F.; Haimson, B.C.; Doe, T.W.; Oldenburg, C.M.; Dobson, P.F. Modeling of hydraulic fracture propagation at the kISMET site using a fully coupled 3D network-flow and quasi-static discrete element model (No. INL/CON-17-41116). In Proceedings of the 42nd Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, CA, USA, 13–15 February 2017.
7. Mistry, D.; Modi, P.; Deokule, K.; Patel, A.; Patki, H.; Abuzaghleh, O. Network traffic measurement and analysis. In Proceedings of the 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, USA, 29 April 2016; pp. 1–7.
8. Olups, R. Zabbix 1.8 Network Monitoring; Packt Publishing Ltd.: Birmingham, UK, 2010.
9. Orazbayev, B.; Ospanov, Y.; Orazbayeva, K.; Makhatova, V.; Kurmangaziyeva, L.; Utenova, B.; Mailybayeva, A.; Mukatayev, N.; Toleuov, T.; Tukpatova, A. System Concept for Modelling of Technological Systems and Decision Making in Their Management; PC Technology Center: Kharkiv, Ukraine, 2021; 180p.
10. Sansyzbay, L.Z.; Orazbayev, B.B. Modeling the operation of climate control system in premises based on fuzzy controller. J. Phys. Conf. Ser. 2019, 1399, 044017.
11. Shao, J.; Zhao, Z.; Yang, L.; Song, P. Remote Monitoring and Control System Oriented to the Textile Enterprise. In Proceedings of the 2009 Second International Symposium on Knowledge Acquisition and Modeling, Wuhan, China, 30 November–1 December 2009; Volume 3, pp. 151–154.
12. Li, Q.; Yang, Y.; Jiang, P. Remote Monitoring and Maintenance for Equipment and Production Lines on Industrial Internet: A Literature Review. Machines 2022, 11, 12.
13. Yugapriya, M.; Judeson, A.K.J.; Jayanthy, S. Predictive Maintenance of Hydraulic System using Machine Learning Algorithms. In Proceedings of the 2022 International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 16–18 March 2022; pp. 1208–1214.
14. Dsouza, J.; Velan, S. Preventive maintenance for fault detection in transfer nodes using machine learning. In Proceedings of the 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE),

Dubai, United Arab Emirates, 11–12 December 2019; pp. 401–404.

15. Polat, H.; Polat, O.; Cetin, A. Detecting DDoS attacks in software-defined networks through feature selection methods and machine learning models. Sustainability 2020, 12, 1035.
16. Wang, M.; Cui, Y.; Wang, X.; Xiao, S.; Jiang, J. Machine learning for networking: Workflow, advances and opportunities. IEEE Netw. 2017, 32, 92–99.
17. Jinglong, Z.; Changzhan, H.; Xiangming, W.; Jiakun, A.; Chunguang, H.; Jinglin, H. Research on Fault Prediction of Distribution Network Based on Large Data. In MATEC Web of Conferences; EDP Sciences: Ulys, France, 2017; Volume 139, p. 00149.
18. Le, T.; Luo, M.; Zhou, J.; Chan, H.L. Predictive maintenance decision using statistical linear regression and kernel methods. In Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA), Barcelona, Spain, 16–19 September 2014; pp. 1–6.
19. Harrell, F.E. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis; Springer: New York, NY, USA, 2001; Volume 608.
20. Liu, T.; Wang, S.; Wu, S.; Ma, J.; Lu, Y. Predication of wireless communication failure in grid metering automation system based on logistic regression model. In Proceedings of the 2014 China International Conference on Electricity Distribution (CICED), Shenzhen, China, 23–26 September 2014; pp. 894–897.