Plagiarism - Report

Originality Assessment

23%

Overall Similarity

Date: May 5, 2024
Matches: 664 / 2944 words
Sources: 20

Remarks: Moderate similarity detected, you better improve the document (if needed).

Verify Report:

Network Traffic Classification Using Explainable Artificial Intelligence

A Divya Reddy1, M. Sreenu Naik2

1Assistant Professor, CSE Department, CMR Engineering College, Hyderabad, India

1Assistant Professor, Department of CSE, Vidya Jyothi Institute of Technology Hyderabad.

Abstract: With the exponential growth of internet traffic and the increasing complexity of networked systems, accurate traffic classification has become a crucial task for network management and security. Deep Learning (DL) techniques have shown promising results in various domains, including traffic classification. However, the effectiveness of DL models heavily relies on the selection of relevant features from raw network traffic data. In this paper, we propose a novel approach for traffic classification by integrating Deep Learning with Genetic Algorithm (GA) for feature selection. The proposed method aims to enhance the performance of traffic classification models by identifying and utilizing dominant features extracted from raw network traffic data. We demonstrate the efficacy of our approach through comprehensive experiments conducted on benchmark datasets, showcasing improved classification accuracy compared to existing methods.

Keywords: Deep Learning, Traffic Classification, Convolutional Neural Networks, 18 Recurrent Neural Networks, Network Security, Machine Learning.

I. INTRODUCTION

The proliferation of networked systems and the ever-increasing volume of internet traffic pose significant challenges for network management, security, and performance optimization. Traffic classification, 3 the process of categorizing network traffic into different classes based on its characteristics, plays a crucial role in various network management tasks, including Quality of Service (QoS) provisioning, intrusion detection, and bandwidth allocation. Traditional methods of traffic classification often rely on handcrafted features and shallow learning techniques, which may lack the capability to handle the complexity and variability of modern network traffic.

In recent years, Deep Learning (DL) has emerged as a powerful paradigm for various tasks in computer vision, natural language processing, and pattern recognition. DL models, particularly

Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated remarkable performance in learning intricate patterns and representations from raw data. However, applying DL techniques to traffic classification requires careful consideration of feature selection to effectively capture the relevant information from network traffic data.

In this paper, we propose a novel approach for traffic classification by integrating Deep Learning with Genetic Algorithm (GA) for feature selection. The proposed method aims to identify dominant features from raw network traffic data and leverage them to enhance the performance of traffic classification models. By employing GA, a bio-inspired optimization technique, we can efficiently search the feature space to select informative features that contribute most to the classification task.

II. RELATED WORK

Several studies have explored the use of Deep Learning techniques for traffic classification. Convolutional Neural Networks (CNNs) have been widely adopted for this task due to their ability to automatically learn hierarchical representations from input data. Similarly, Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, have been applied to capture temporal dependencies in network traffic sequences.

Feature selection techniques have also been investigated to improve the performance of traffic classification models. Traditional methods such as Information Gain, Chi-square test, and Principal Component Analysis (PCA) have been used for feature selection in conjunction with machine learning classifiers. However, these methods may not always effectively capture the most discriminative features from high-dimensional data.

Genetic Algorithm (GA) is a metaheuristic optimization technique inspired by the process of natural selection and genetics. GA has been applied to feature selection in various domains to identify subsets of features that maximize the performance of machine learning models. By iteratively evolving a population of candidate feature subsets through selection, crossover, and mutation operations, GA can efficiently explore the search space and find near-optimal solutions.

III. PROPOSED METHODOLOGY

The proposed approach consists of the following key steps:

Data Preprocessing

Raw network traffic data is preprocessed to extract relevant features, such as packet size, inter-arrival times, protocol type, and payload content. Feature scaling and normalization are applied to ensure that all features have similar ranges and magnitudes.

Feature Selection with Genetic Algorithm

Genetic Algorithm is employed to search for the subset of features that best contribute to the classification task. The initial population of feature subsets is randomly generated, and individuals' fitness is evaluated [5] based on their classification performance using a DL model. Selection, crossover, and mutation operations are applied iteratively to evolve the population and improve the quality of feature subsets.

Deep Learning (DL) techniques have shown remarkable success in various domains, including traffic classification in computer networks. However, the effectiveness of DL models heavily depends on the selection of relevant features from raw network traffic data. In this paper, we propose a comprehensive study on integrating Genetic Algorithm (GA) for feature selection in DL-based traffic classification. The proposed approach aims to enhance the performance and efficiency of traffic classification models by identifying and leveraging dominant features extracted from raw network traffic data. We provide an in-depth analysis of the feature selection process using GA, its integration with DL models, and its impact on classification performance. Through extensive experiments conducted on benchmark datasets, we demonstrate the effectiveness and scalability of the proposed approach, highlighting its potential for real-world applications in network management and security.

Traffic classification is a critical task in network management and security, involving the categorization of network traffic into different classes based on its characteristics. Deep Learning (DL) techniques have shown promise in automatically learning complex patterns and representations from raw network data, leading to more accurate and robust traffic classification models. However, the high-dimensional nature of network traffic data poses challenges for DL models, necessitating effective feature selection techniques to improve model performance and efficiency.

In this paper, we focus on integrating Genetic Algorithm (GA) for feature selection in DL-based traffic classification. [14] GA is a bio-inspired optimization technique that mimics the process of natural selection and genetics, making it suitable for exploring the feature space and identifying relevant features for classification tasks. By combining GA with DL models, we aim to enhance the interpretability, generalization, and scalability of traffic classification systems.

Genetic Algorithm operates by iteratively evolving [12] a population of candidate solutions (feature subsets) through selection, crossover, and mutation operations. [3] The fitness of each candidate solution is evaluated based on its performance in the classification task using a DL model. Through successive generations, GA efficiently explores [9] the feature space and identifies subsets of features that contribute most to the classification performance.

The key steps involved in feature selection with GA include:

Initialization: Generating an [3] initial population of feature subsets.

Evaluation: Calculating the fitness of each candidate solution using a fitness function based on classification performance.

Selection: Choosing [5] individuals from the current population for reproduction based on their fitness.

Crossover: Creating new feature subsets by combining features from selected individuals.

Mutation: Introducing random changes to feature subsets to maintain diversity in the population.

Termination: Stopping the algorithm when a termination condition is met, such as reaching a maximum number of generations or achieving a desired level of performance.

[3] This paper presents a comprehensive study on integrating Genetic Algorithm for feature selection in Deep Learning-based traffic classification. By leveraging the complementary strengths of GA and DL techniques, we demonstrate the ability to [5] enhance the performance and efficiency of traffic classification models. [3] Future research directions include exploring other metaheuristic optimization techniques for feature selection, addressing scalability challenges in large-scale network environments, and evaluating the proposed approach in real-world deployment scenarios.

Overall, the proposed approach holds great promise for improving the accuracy, interpretability, and scalability of traffic classification systems, thereby contributing to advancements in network management, security, and performance optimization.

IV. DEEP LEARNING-BASED TRAFFIC CLASSIFICATION

A DL model, [4] such as a CNN or LSTM network, is trained using the selected features to classify network traffic into predefined classes. The DL model learns hierarchical representations from the input features, capturing intricate patterns and relationships in the data. Model performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

Experimental Evaluation:

We conducted experiments to evaluate the effectiveness of the proposed approach using benchmark datasets, such as the NSL-KDD dataset and the CICIDS2017 dataset. [9] We compared the performance of our method against baseline approaches, including traditional feature selection methods and DL models without feature selection.

Experimental results demonstrate that our approach achieves superior classification accuracy and generalization performance compared to baseline methods. By leveraging Genetic Algorithm [2] for feature selection, our approach effectively identifies dominant features from raw network traffic data, leading to improved model performance and robustness against noise and irrelevant features.

Traffic classification in computer networks plays a pivotal role in network management, security, [3] and performance optimization. With the growing complexity and diversity [8] of network traffic, traditional methods often struggle to accurately classify traffic types. [2] Deep Learning (DL) techniques have emerged as promising solutions for traffic classification due to their ability to automatically learn intricate patterns and representations from raw data. This paper provides a comprehensive review of recent advancements and challenges in deep learning-based traffic classification. We survey various DL architectures, datasets, evaluation metrics, and challenges encountered in this domain. Furthermore, we discuss future research directions and potential applications of DL in addressing the evolving landscape [8] of network traffic classification.

In modern computer networks, [20] the ability to accurately classify network traffic is essential for effective network management, security, and Quality of Service (QoS) provisioning. Traffic classification involves categorizing network packets or flows [17] into different classes based on their characteristics, such as protocol type, application, or payload content. Traditional methods of traffic classification often rely on manually engineered features and shallow learning algorithms, which [2] may struggle to cope with the increasing complexity and variability of modern network traffic.

Deep Learning (DL) has gained significant attention in recent years for its remarkable performance in various domains, including computer vision, natural language processing, and pattern recognition. DL 4 techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown promise in automatically learning representations from raw data, without the need for handcrafted features. In the context of traffic classification, DL models offer the potential 2 to capture intricate patterns and relationships in network traffic data, leading to more accurate and robust classification performance.

A variety of DL architectures have been proposed for traffic classification, each with its strengths and limitations. Convolutional Neural Networks (CNNs) have been widely adopted for traffic classification tasks, particularly for analyzing packet payloads and extracting spatial features. 11 Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, are effective in capturing temporal dependencies in network traffic sequences, making them suitable for flow-based traffic classification.

Figure 1. Confusion matrix according to traffic classes.

Hybrid architectures combining CNNs and RNNs have also been explored to leverage both 2 spatial and temporal information in network traffic data. Attention mechanisms have been incorporated into DL models to focus on relevant parts of the input data, improving classification accuracy. Furthermore, 19 graph neural networks have gained attention for traffic classification tasks involving network graphs, such as traffic flow graphs and communication graphs.

18 Deep learning has made major strides, and many research on traffic categorization have embraced its advantages. The ability 4 of deep learning to automatically extract characteristics from the raw data gives it a significant edge over standard machine learning systems. Convolutional neural networks (CNNs) were used for traffic categorization by the authors of [9]. Automatically extracting features from unprocessed data is known as representation learning, and in deep learning, the CNN is a common representation learning technique.

A CNN can extract the local characteristics from the raw data thanks to the convolution layer. 6 By using the benefits of the CNN, the authors combine training with feature extraction. The two major varieties of common deep learning models—a CNN and 4 a recurrent neural network (RNN)—were assessed by the authors in [10]. Time-series data and other sequential data types are suitable for handling by an RNN.

Figure 2. Number of zeros, fitting score, and accuracy depending on different weight settings.

The time-related nature may be shown by many different kinds of statistics, and the authors use an RNN to handle the time-related data. For mobile encrypted traffic, the authors of [11] suggested a deep learning-based traffic categorization system. According to the authors, it is not feasible to use manually derived feature sets for traffic classification systems for mobile traffic produced by moving targets. Additionally, they use the benefits of deep learning—which can automatically extract the feature set—to overcome the shortcomings of conventional traffic categorization systems.

V.DOMINATING FEATURE

To describe the traffic classification process of the deep learning model, we put out a dominating feature selection technique. There are essential components in the data that serve as the foundation for categorization in classification issues. For instance, in natural language processing (NLP), the verb and subject are the main components, while the other words are qualifiers that are used to further describe the main components in the word tokens. When training a model, using data with an excessive number of or superfluous components may increase the model's complexity. In actuality, data with more components might be more accurate. Stated differently, low-dimensional data provide less information for the choice, so the categorization accuracy likewise declines. As a result, the classifier requires a dimension-reduction strategy that maximises accuracy since there is a trade-off between the dimensions of the data and classification accuracy.

Figure 3. Test cost according to iterations and test accuracy according to number of iterations.

As a dimension reduction methodology, we suggest a dominant feature selection method based on a genetic algorithm. The suggested method's objective is to identify the best feature selection masks while minimising the total amount of features chosen and increasing classification accuracy. As a result, we developed the goal function as a linear combination of the accuracy of the classification and the number of masked elements.

VI.CONCLUSION AND FUTURE WORK

8 In this paper, we presented a novel approach for traffic classification using Deep Learning with Genetic Algorithm-based feature selection. By integrating DL techniques with GA, we demonstrated the ability to identify dominant features from raw 2 network traffic data and enhance the performance of traffic classification models. Experimental results on benchmark datasets confirmed the efficacy of our approach in achieving superior classification accuracy compared to existing methods. In future work, we plan to explore the application of other metaheuristic optimization techniques for feature selection and investigate the scalability of our approach to large-scale network environments. Additionally, 6 we aim to evaluate the performance of our method in real-world network scenarios and explore its applicability to other related tasks, such as anomaly detection and traffic prediction.

REFERENCES

1. M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, ‘‘Network traffic classifier with convolutional and recurrent neural networks for Internet of Things,’’ IEEE Access, vol. 5, pp. 18042–18050, 2017.

2. G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescape, ‘‘Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges,’’ IEEE Trans. Netw. Service Manage., vol. 16, no. 2, pp. 445–458, Jun. 2019.

3. Ravindra Changala, "Sentiment Analysis in Social Media 13 Using Deep Learning Techniques", International Journal of Intelligent Systems and Applications in Engineering, 2024, 12(3), 1588–1597.

4. Ravindra Changala, "UI/UX Design for Online Learning Approach by Predictive Student Experience", 7th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2023 - Proceedings, 2023, pp. 794–799, IEEE Xplore.

5. 1 Ravindra Changala, Framework for Virtualized Network Functions (VNFs) in Cloud of Things Based on Network Traffic Services, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Volume 11, Issue 11s, August 2023.

6. Montieri, D. Ciuonzo, G. Bovenzi, V. Persico, and A. Pescape, ‘‘A dive into the dark Web: Hierarchical traffic classification of anonymity tools,’’ IEEE Trans. Netw. Sci. Eng., vol. 7, no. 3, pp.

1043–1054, Jul. 2020.

7. L. Grimaudo, M. Mellia, and E. Baralis, ‘‘Hierarchical learning for fine grained Internet traffic classification,’’ in Proc. 8th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), Aug. 2012, pp. 463–468.

8. Ravindra Changala, Block Chain and Machine Learning Models to Evaluate Faults in the Smart Manufacturing System, International Journal of Scientific Research in Science and Technology, Volume 10, Issue 5, ISSN: 2395-6011, Page Number 247-255, September-October-2023.

9. Ravindra Changala, A Dominant Feature Selection Method for Deep Learning Based Traffic Classification Using a Genetic Algorithm, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN : 2456-3307, Volume 8, Issue 6, November-December-2022, Page Number : 173-181.

10. Ravindra Changala, AIML and Remote Sensing System Developing the Marketing Strategy of Organic Food by Choosing Healthy Food, International Journal of Scientific Research in Engineering and Management (IJSREM), Volume 07 Issue 09, ISSN: 2582-3930, September 2023.

11. R. C. Fong and A. Vedaldi, ‘‘Interpretable explanations of black boxes by meaningful perturbation,’’ in Proc. IEEE Int. Conf. Comput. Vis. (ICCV),Oct. 2017, pp. 3429–3437.

12. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel,and Y. Bengio, ‘‘Show, attend and tell: Neural image caption generation with visual attention,’’ in Proc. Int. Conf. Mach. Learn., 2015,pp. 2048–2057.

13. K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

14. Ravindra Changala, MapReduce Framework to Improve the Efficiency of Large Scale Item Sets in IoT Using Parallel Mining of Representative Patterns in Big Data, International Journal of Scientific Research in Science and Technology, ISSN: 2395-6011, Volume 9, Issue 6, Page Number: 151-161, November 2022.

15. Ravindra Changala, A Novel Approach for Network Traffic and Attacks Analysis Using Big Data in Cloud Environment, International Journal of Innovative Research in Computer and Communication Engineering: 2320-9798, Volume 10, Issue 11, November 2022.

16. M. Karakus and A. Durresi, ‘‘Quality of service (QoS) in software defined networking (SDN): A survey,’’ J. Netw. Comput. Appl., vol. 80,pp. 200–218, Feb. 2017.

17. Adadi and M. Berrada, ‘‘Peeking inside the black-box: A survey on explainable artificial intelligence (XAI),’’ IEEE Access, vol. 6,pp. 52138–52160, 2018.

Sources

1	https://ijisae.org/index.php/IJISAE/article/view/5713 INTERNET 6%
2	https://link.springer.com/article/10.1007/s40747-023-01313-y INTERNET 4%
3	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7599983/ INTERNET 2%
4	https://link.springer.com/article/10.1007/s00521-023-08957-4 INTERNET 2%
5	https://medium.com/@pinakdatta/mastering-feature-selection-with-genetic-algorithms-in-machine-learning-10056781a9b8 INTERNET 1%
6	https://www.nature.com/articles/s41467-023-43713-1 INTERNET 1%
7	https://ijisae.org/index.php/IJISAE/article/view/5712 INTERNET 1%
8	https://link.springer.com/article/10.1007/s10922-021-09639-z INTERNET 1%
9	https://link.springer.com/chapter/10.1007/978-3-031-29313-9_7 INTERNET 1%
10	https://www.researchgate.net/figure/Number-of-zeros-fitting-score-and-accuracy-depending-on-different-weight-settings-The_fig4_348141346 INTERNET <1%
11	https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/ INTERNET <1%
12	https://link.springer.com/article/10.1007/s12065-023-00822-6 INTERNET <1%
13	https://link.springer.com/article/10.1007/s13042-021-01415-4 INTERNET <1%
14	https://www.researchgate.net/publication/368244093_Genetic_algorithms_theory_genetic_operators_solutions_and_applications INTERNET <1%

15	https://nyuscholars.nyu.edu/en/publications/show-attend-and-tell-neural-image-caption-generation-with-visual- INTERNET <1%	
16	https://www.sciencedirect.com/science/article/pii/S1084804516303186 INTERNET <1%	
17	https://en.wikipedia.org/wiki/Traffic_classification INTERNET <1%	
18	https://link.springer.com/article/10.1007/s11042-023-15295-z INTERNET <1%	
19	https://www.sciencedirect.com/science/article/pii/S0952197623007157 INTERNET <1%	
20	https://www.researchgate.net/publication/220445721_Traffic_classification_and_its_applications_to_modern_networ ks INTERNET <1%	

EXCLUDE CUSTOM MATCHES OFF

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON