Faculty of Computer Science and Information Technology

Enhanced Adaptive Neuro-Fuzzy Inference System Classification Method for Intrusion Detection

Liu Jia

Doctor of Philosophy
2024

Enhanced Adaptive Neuro-Fuzzy Inference System Classification Method for Intrusion Detection

Liu Jia

A thesis submitted

In fulfillment of the requirements for the degree of Doctor of Philosophy

(Software Engineering)

Faculty of Computer Science and Information Technology
UNIVERSITI MALAYSIA SARAWAK
2024

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

……

Signature 

Name: Liu Jia

Matric No.: 20010028

Faculty of Computer Science and Information Technology

Universiti Malaysia Sarawak

Date :

ACKNOWLEDGEMENT

I would like to take this opportunity to those who have contributed directly or indirectly to this research.

My sincere gratitude to the Faculty of Computer Science and Information Technology, for the advice and support given during my period of study in Universiti Malaysia Sarawak.

I would like to thank my two supervisors, Professor Dr. Wang Yin Chai and Associate Professor Dr. Teh Chee Siong, for their guidance. In the process of learning, I benefited a lot from their serious attitude and spirit of learning.

I am also grateful to my parents and husband for their support during my studies. Without them, I would not have been able to complete my studies on time.

Finally, I would like to thank the management of the Universiti Malaysia Sarawak for making it possible for me to complete my study here in Sarawak. Thank you all.

ABSTRACT

Nowadays, the rapid advancement of technologies such as e-commerce, mobile payments, cloud computing, big data and analytics, IoT, AI, machine learning, and social media has greatly contributed to economic growth. However, it has also resulted in an increased frequency of cyberattacks. In order to ensure network security, intrusion detection has become a crucial technology. However, due to the inherent uncertainty in distinguishing between normal and abnormal behaviours, intrusion detection problems can be considered as fuzzy classification problems. Many traditional classification methods are inadequate in handling this fuzziness, leading to suboptimal performance in intrusion detection. To address this issue, the Adaptive-Network-based Fuzzy Inference System (ANFIS) has emerged as a promising approach. ANFIS combines the uncertainty processing ability of fuzzy logic with the learning process of Neural Networks (NNs). It has demonstrated excellent performance in various domains, including intrusion detection. However, due to the constraint of having only 5 layers, ANFIS is unable to identify higher-level and more abstract representations of the data. To address this problem, this study first uses CART (Classification and Regression Tree) to enhance the depth of ANFIS, providing a deeper and interpretable hybrid architecture. The method is then compared to single CART, single ANFIS, BPNN, GA-ANFIS, PSO-ANFIS, and CSO-ANFIS. The results demonstrate a detection rate of 99.86% and a false alarm rate of 0.14% on the intrusion detection dataset, KDDTrain+, which is one of the most widely used benchmark datasets. The proposed method outperforms CSO-ANFIS by exhibiting a 4.06% higher detection rate (improving from 95.80%) and a 2.31% lower false alarm rate (compared to 3.45%). On the KDDTest+ dataset, the proposed method also outperforms single CART and ANFIS in terms of various metrics other than precision. Since the CART tree is a binary tree, it can only represent the relationship between data

through a split based on a single attribute at a single tree node. Therefore, this binary tree cannot analyse complex features of mixed attributes and restricts the CART tree's deep-level feature recognition ability. Although the combination of ANFIS and CART provides deeper feature recognition for ANFIS, it still lacks the ability to recognize deep-level features and mixed-attribute features due to the limitations of both ANFIS and CART. Therefore, deep-level feature analysis, recognizing complex and mixed factors, has essential research value and significance for improving the efficiency and accuracy of intrusion detection. This study further enhances the depth of ANFIS and identifies deep features by using the ResNet (Residual Network) with ANFIS. The proposed hybrid model is able to deepen the structure of ANFIS and provide a deep and hybrid architecture. A comparison study with the proposed method with MVO-ANN, FC-ANN, DN, CSO-ANFIS, single ANFIS and single ResNet is conducted. The results show improved performance with a 10.68% better detection rate (improving from 88.2%) and a 10.68% lower false alarm rate (compared to 11.79%) than a single ANFIS. Furthermore, the proposed method outperforms MVO-ANN, FC-ANN, DN, CSO-ANFIS, single ANFIS, and single ResNet in terms of detection rate, F1-score, and recall rate. Additionally, standard deviation and proposed adaptive K-means algorithms have been employed to minimize the generated rules by ANFIS from the proposed hybrid models. This approach can dynamically minimize the interval number of each continuous attribute. As a result, the number of fuzzy rules generated by ANFIS is reduced, subsequently improving the training and prediction efficiency of the proposed hybrid models.

Keywords: Adaptive Neuro-fuzzy Inference System, Interpretable Hybrid Neural Networks, Classification and Regression Tree, Deep Residual Networks, Intrusion detection

Kaedah Klasifikasi Sistem Inferens Neuro-Fuzzy Adaptive yang Dipertingkatkan untuk Pengesanan Pencerobohan

ABSTRAK

Dalam zaman moden ini, kemajuan pesat dalam teknologi seperti e-dagang, pembayaran mudah alih, pengkomputeran awan, data besar dan analitik, IoT, AI, pembelajaran mesin, dan media sosial telah memberikan sumbangan yang besar kepada pertumbuhan ekonomi. Walau bagaimanapun, ia juga telah mengakibatkan peningkatan frekuensi serangan siber. Bagi memastikan keselamatan rangkaian, pengesanan pencerobohan telah menjadi teknologi yang penting. Namun demikian, disebabkan ketidakpastian dalam membezakan tingkah laku normal dan tidak normal, masalah pengesanan pencerobohan boleh dianggap sebagai masalah klasifikasi yang kabur. Banyak kaedah klasifikasi tradisional tidak memadai dalam menangani ketidakpastian ini, yang membawa kepada prestasi yang kurang memuaskan dalam pengesanan pencerobohan. Bagi menangani isu ini, Sistem Inferens Kabur Berasaskan Rangkaian Adaptif (ANFIS) adalah pendekatan yang sesuai dipilih. ANFIS menggabungkan keupayaan pemprosesan ketidakpastian logik kabur dengan proses pembelajaran Rangkaian Neural (NNs). Ia telah menunjukkan prestasi klasifikasi yang baik dalam pelbagai bidang, termasuk pengesanan pencerobohan. Walau bagaimanapun, disebabkan sistem ini hanya mempunyai 5 lapisan NNs, kekangan ini menyebabkan ANFIS tidak dapat mengenal perwakilan data yang rumit dan lebih abstrak. Untuk mengatasi masalah ini, kajian ini terlebih dahulu menggunakan CART (Classification and Regression Tree) untuk meningkatkan kedalaman ANFIS. Rekabentuk hibrid yang berlapisan lebih dalam ini akan berupaya menerang perwakilan data yang rumit dan lebih abstrak. Kaedah ini kemudiannya dibandingkan dengan single CART, single ANFIS, BPNN, GA-ANFIS, PSO-ANFIS, dan CSO-ANFIS. Keputusan menunjukkan kadar pengesanan sebanyak

99.86% dan kadar penggera palsu sebanyak 0.14% diperolehi dengan menggunakan set data pengesanan pencerobohan (KDDTrain+), yang mana set data ini merupakan salah satu set data penanda aras yang paling banyak digunakan. Kaedah yang dicadangkan dapat meningkatkan kadar pengesanan 4.06% lebih tinggi (bertambah baik daripada 95.80%) dan kadar penggera palsu 2.31% lebih rendah (berbanding 3.45%) dengan model CSO-ANFIS. Dengan menggunakan dataset KDDTest+, kaedah yang dicadangkan juga mengatasi prestasi CART tunggal dan ANFIS dari pelbagai ukuran metrik. Disebabkan pepohon CART adalah pepohon binari, ia hanya dapat mewakili hubungan data melalui atribut tunggal pada satu node pepohon. Oleh itu, pepohon binari ini tidak dapat menganalisa sifat data yang mempunyai atribut campuran yang kompleks, ini selanjutnya membatasi kemampuan pengenalan sifat data dengan pepohon daripada CART. Walaupun kombinasi ANFIS dan CART menghasilkan lapisan NNs yang lebih dalam untuk ANFIS, ia masih kurang mampu mengenali sifat data yang mendalam dan atribut data yang bersifat campuran kerana limitasi oleh kedua-dua model ANFIS dan CART. Oleh kerana itu, satu model yang dapat meningkatkan efisiensi dan ketepatan menganalisa sifat data yang mendalam dan atribut data yang bersifat campuran dan kompleks perlu dihasilkan untuk pengesanan pencerobohan. Dalam kajian ini, kedalaman lapisan ANFIS akan ditingkatkan dengan menggunakan ResNet (Residual Network). Model hybrid yang dicadangkan dalam kajian ini mampu memperdalamkan struktur ANFIS. Beberapa kajian perbandingan telah dilaksanakan dengan model hybrid yang dicadangkan dan model-model lain, seperti MVO-ANN, FC-ANN, DN, CSO-ANFIS, ANFIS tunggal dan ResNet tunggal. Keputusan menunjukkan prestasi perningkatan kadar pengesanan sebanyak 10.68% (bertambah baik daripada 88.2%) dan kadar penggera palsu 10.68% lebih rendah (berbanding 11.79%) dengan model ANFI tunggal. Kaedah yang dicadangkan juga mengatasi prestasi MVO-

ANN, FC-ANN, DN, CSO-ANFIS, ANFI tunggal, dan ResNet tunggal dari segi kadar pengesanan, skor F1, dan kadar recall. Selain itu, sisihan piawai dan algoritma k-means adaptif yang dicadangkan juga dapat mengurangkan tanda peraturan (rules) yang dihasilkan oleh ANFIS daripada model hibrid yang dicadangkan. Pendekatan dinamik ini dapat meminimumkan bilangan selang setiap atribut berterusan data yang digunakan. Akibatnya, bilangan tanda peraturan kabur yang dihasilkan oleh ANFIS juga dikurangkan, dan selanjutnya meningkatkan kecekapan latihan dan keupayaan ramalan model hibrid yang dicadangkan. Pendekatan ini secara dinamik dapat meminimumkan bilangan selang setiap atribut berterusan. Akibatnya, bilangan peraturan kabur yang dihasilkan oleh ANFIS dikurangkan, seterusnya meningkatkan kecekapan latihan dan ramalan model hibrid yang dicadangkan.

Kata kunci: *Sistem Inferensi Neuro-Fuzzy Adaptif, Rangkaian Neural Hybrid Boleh Tafsir, Pepohon Klasifikasi dan Regresi, Jaringan Residual Dalam, Pengesanan Pencerobohan*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ABC	Artificial Bee Colony
ACC	Accuracy
AI	Artificial Intelligence
ANFIS	Adaptive Network-based Fuzzy Inference System
ANN	Artificial Neural Network
BP	Back Propagation
BPNN	Back Propagation Neural Network
CART	Classification and Regression Tree
CGS	Centre for Graduate Studies
CNCERT	National Internet Emergency Centre of China
CNN	Convolutional Neural network
DBN	Deep Belief Network
DDoS	Distributed Denial of Service
DNN	Deep Neural Network
DOS	Denial-Of-Service
DT	Decision Tree
FAR	False Alarm Rate
FN	False Negative
FP	False Positive
GA	Generic Algorithm
GAN	Generative Adversarial Network
GBDT	Gradient Boosting Decision Tree
GNN	Graph Neural Network

GoogleNet	GoogleInceptionNet
HIDS	Host-based IDS
HMM	Hidden Markov Model
IDS	Intrusion Detection System
IoT	Internet of Things
K-means	K-means Clustering Algorithm
KNN	K-nearest Neighbour
LAN	Local Area Network
LSTM	Long Short-Term Memory
MF	Membership Function
NIDS	Network-based IDS
NN	Neural Network
NSL-KDD	National Security Lab–Knowledge Discovery and Data Mining
PCA	Principal component analysis
PROBE	Surveillance or Probe
R2L	Remote to Local
RBM	Restricted Boltzmann Machine
ResNet	Residual Network
RNN	Recurrent Neural Network
SAE	Stacked Autoencoder
SOM	Self-Organizing Map
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
U2R	User to Root
UNIMAS	Universiti Malaysia Sarawak

VGG	Visual Geometry Group
WDLSTM	Weight-Dropped Long Short-term Memory
XGBoost	Extreme Gradient Boosting

CHAPTER 1

INTRODUCTION

1.1 Background

The security of the network is becoming increasingly precarious. The Cisco Annual Internet Report (2018–2023) White Paper (Cisco, 2020) reveals a concerning trend of increasing network intrusions. Specifically, there has been a massive 776% surge in attacks with data volumes ranging from 100 Gbps to 400 Gbps between 2018 and 2019. This indicates a clear upward trend in the frequency and severity of cyberattacks (Cisco, 2020).

Furthermore, the report indicates that more than half of the surveyed network operators have encountered infrastructure outages. This highlights the significant impact of cyberattacks on the stability and reliability of network infrastructures, posing serious challenges to the cybersecurity landscape. These outages can have detrimental effects on businesses and individuals, leading to financial losses, service disruptions, and compromised data security.

Given the escalating threat landscape, it is crucial for organizations to prioritize network security measures and implement robust intrusion detection systems. Proactive monitoring, timely threat intelligence, and continuous security updates are vital to safeguarding networks from potential breaches and minimizing the impact of cyberattacks.

In China, the network security scenario is also of great concern. The 50th Statistical Report on China's Internet Development (CINIC, 2022) reveals that the National Internet Emergency Centre of China (CNCERT) received 15,654 reports of network security

incidents in the first half of 2022. Furthermore, China Telecom, China Mobile, and China Unicom detected a total of 16,542 DDoS attacks in the same period.

The rapid progress of technology has presented us with numerous opportunities, but it has also posed significant challenges in the domain of network security. The emergence and widespread adoption of technologies like AI (Artificial Intelligence), e-commerce, social media, machine learning, IoT (Internet of Things), cloud computing, Big Data, and mobile payments have undoubtedly fuelled economic growth. However, alongside these advancements, there has been a corresponding rise in the number and sophistication of cyberattacks (Cisco, 2020).

Cybercriminals have leveraged the potential vulnerabilities brought about by these technologies to launch attacks that are increasingly covert and intelligent. This has rendered traditional intrusion detection methods less effective in identifying and mitigating such attacks. Given the increasing complexity and sophistication of cyber threats, there is an urgent requirement to develop and implement more advanced and adaptive security measures to protect our networks and sensitive data (imarc, 2022).

Consequently, the study of new techniques for intrusion detection has become of paramount importance. Intrusion detection, being a crucial component of network security, has assumed a more critical role in the evolving network landscape. It is imperative that further research and development be carried out in this field to ensure the security of networks against potential threats.

However, intrusion detection is a fuzzy classification problem that involves uncertainty in the classification process. The behaviour of intruders in network security often exhibits similarities to that of legitimate users, as illustrated in Figure 1.1. This overlap

results in inherent fuzziness and uncertainty in intrusion detection. Many traditional machine learning methods are not well equipped to handle this uncertainty (Tsang et al., 2007).

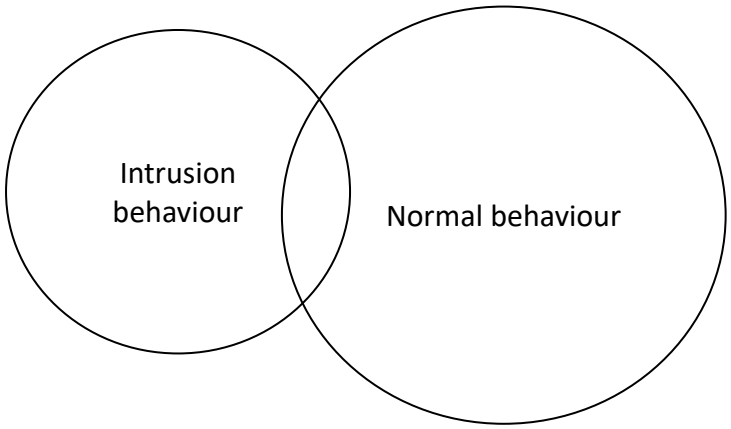

Figure 1.1: Intrusion Behaviours and Normal Behaviours

Since the introduction of fuzzy logic theory by Lotfi Zadeh in 1965, it has been regarded as a powerful approach to address problems involving uncertainty (Masdari & Khezri, 2020). Fuzzy logic is a method that uses fuzzy sets to represent linguistic variables and allows for the specification of intermediate values. Fuzzy logic-based techniques have gained significant popularity and widespread usage in the development of intrusion detection systems., such as fuzzy clustering, fuzzy classifiers, fuzzy feature extraction, and others. The effectiveness of these methods in detecting intrusions in the presence of uncertain data has been demonstrated by researchers (Masdari & Khezri, 2020). While fuzzy logic provides a powerful framework for modelling and handling uncertainty, it is still limited by its inability to fully capture the complex and dynamic nature of the intrusion detection problem. This highlights the need for further research to improve the application of fuzzy logic in intrusion detection, to enhance its ability to handle the uncertainties and ambiguities of the problem and ultimately improve its performance (Alsaadi et al., 2022).

Recently, Neural Network (NN) algorithms have gained considerable attention due to their superior performance in image classification tasks, leading to the development of various NN-based intrusion detection systems (Drewek-Ossowicka et al., 2021). These systems leverage the deep feature extraction abilities of NNs, enabling them to cope with the complex and dynamic network environment. However, the issue of overfitting, which occurs when the model is too complex and fits the training data too closely, resulting in poor generalization performance on unseen data, remains prevalent in NN-based intrusion detection systems applied to fuzzy classification problems. In particular, overfitting can cause the model to overlook important patterns in the data, which are crucial for accurate classification, making it a challenging problem that must be addressed to ensure robust and effective performance (Drewek-Ossowicka et al., 2021).

To address this issue, scholars have proposed various solutions. Some of the most popular methods include regularization techniques such as dropout and early stopping, as well as model architecture modifications such as pruning and ensemble methods (Piotrowski et al., 2020). However, intrusion detection is a fuzzy classification problem. To better fit the fuzzy nature of this problem, the combination of fuzzy logic and neural networks is an inevitable choice (Alsaadi et al., 2022).

Fuzzy schemes have proven to be effective in detecting intrusions and malicious behaviors when dealing with uncertain data (Masdari & Khezri, 2020). Fuzzy logic has emerged as a popular solution for addressing uncertainty in intrusion detection and has demonstrated its effectiveness in this field. Among the different fuzzy approaches applied in intrusion detection, the ANFIS classifier has gained widespread adoption (Masdari & Khezri, 2020).

By combining the benefits of fuzzy logic and neural networks, Adaptive Neuro-fuzzy Inference System (ANFIS) is designed to provide an efficient solution to the problems of uncertainty in intrusion detection (Jang, 1993). On the KDDCUP99 dataset, the use of ANFIS in intrusion detection systems was first reported by Toosi and Kahani (Toosi & Kahani, 2007), who achieved a 95.3% detection rate on the KDDCUP99 dataset using five ANFIS modules. While ANFIS demonstrates high performance in fuzzy classification tasks, it does have certain limitations that warrant further investigation in future research. One of the key limitations is the lack of deep feature recognition capability (T.-L. Nguyen, 2023).

1.2 Problem Statement

Intrusion detection is a security measure designed to identify intrusion behaviors. These detection problems are characterized by a degree of uncertainty in the classification process. Often, the behaviors of intruders overlap with those of legitimate users, leading to inherent uncertainty in intrusion detection. This uncertainty poses a significant challenge for many traditional machine learning methods, as they may struggle to accurately differentiate between intruder and legitimate user behaviors in the presence of such overlap (Tsang et al., 2007).

It is crucial to develop algorithms that can effectively handle this uncertainty in intrusion detection. Fuzzy logic has been shown to be a promising approach (Masdari & Khezri, 2020). In various fuzzy-based approaches, ANFIS is one of the most effective and widely used algorithms for intrusion detection (Masdari & Khezri, 2020).

ANFIS, which stands for Adaptive-Network-based Fuzzy Inference System, is a method that combines fuzzy logic and neural network models to create a powerful tool for intrusion detection. ANFIS's capability to generate fuzzy and overlapping fuzzy rules,