A NEURAL NETWORK MODAL DECOMPOSITION MECHANISM IN PREDICTING NETWORK TRAFFIC

SHI JINMEI

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH

2023

A NEURAL NETWORK MODAL DECOMPOSITION MECHANISM IN PREDICTING NETWORK TRAFFIC

SHI JINMEI

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH
2023

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL : **A NEURAL NETWORK MODAL DECOMPOSITION MECHANISM IN PREDICTING NETWORK TRAFFIC**

IJAZAH : **DOKTOR FALSAFAH SAINS KOMPUTER**

BIDANG : **SAINS KOMPUTER**

Saya **SHI JINMEI**, Sesi **2019-2023**, mengaku membenarkan tesis Doktoral ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:

1. Tesis ini adalah hak milik Universiti Malaysia Sabah
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/):

[ ]	SULIT	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA 1972)
[ ]	TERHAD	(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
[/]	TIDAK TERHAD	

Disahkan Oleh,

ANITA BINTI ARSAD
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH

SHI JINMEI

SHI JINMEI
DI1821011A

(Tandatangan Pustakawan)

DR. LEAU YU BENG
Faculty of Computing and Informatics
Universiti Malaysia Sabah
lybeng@ums.edu.my

Dr. Leau Yu Beng

Tarikh : 25 July 2023

(Dr. Leau Yu Beng)

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries, and references, which have been duly acknowledged.

11 April 2023

SHI JINMEI

Shi Jinmei
DI1821011A

CERTIFICATION

NAME : **SHI JINMEI**

MATRIC NUM. : **DI1821011A**

TITLE : **A NEURAL NETWORK MODAL DECOMPOSITION MECHANISM IN PREDICTING NETWORK TRAFFIC**

DEGREE : **DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE**

FIELD : **COMPUTER SCIENCE**

VIVA DATE : **11 APRIL 2023**

CERTIFIED BY;

Signature

1. MAIN SUPERVISOR
Dr. Leau Yu Beng

DR. LEAU YU BENG
Faculty of Computing and Informatics
lybeng@ums.edu.my

Dr. Leau Yu Beng

2. CO-SUPERVISOR
Prof. Dr. Li Kun

Kunli

ACKNOWLEDGEMENT

I would like to take this opportunity to express my deepest gratitude to my supervisors Dr. Leau Yu Beng and Associate Prof. Dr. Li Kun for their invaluable guidance, patience, and encouragement until the end of my academic journey. In particular, Dr. Leau's guidance to me made me feel warm as a Chinese student and also motivated me to continue my Ph.D. career.

I would like to thank the researchers in the field of network traffic prediction for providing us with academic achievements on time series prediction models, which can be used for reference to help us complete this research.

At the same time, I would like to thank the President of my university Dr. Yang Xiuying, who gave me full tuition support and gave me the opportunity to pursue my Ph.D. in the faculty of computing and informatics Universiti Sabah Malaysia.

Finally, I would like to thank my dear husband for his love, understanding, support, and encouragement until the completion of my academic journey.

Shi Jinmei
11 April 2023

ABSTRACT

Network traffic prediction is essential for effective network management as it can provide an early warning to the administrator before an incident occurs. This study designs a novel network traffic prediction model namely SAVE-AS. It embeds a new proposed Scalable Artificial Bee Colony (SABC) algorithm, Phase Space Reconstruction, Variational Mode Decomposition (VMD) and an integrated Extreme Learning Machine (ELM). The proposed mechanism starts by using SABC to update the model with a new solution and fine-tune the disturbances in each iteration to deal with the interference in order to find the best values that are also synchronously optimal. The SAVE-AS then constructs an adaptive selection operator. It adaptively selects the number of datasets after VMD optimization decomposition to precisely set the number of hidden layer nodes in an ELM to improve prediction accuracy. Meanwhile, the ELM model is trained using a variety of sub-data sequences that meet the requirements for minimizing computational complexity in modeling. Furthermore, the mechanism eliminates the poor sub-sequence caused by the volatility of the results to accelerate the convergence rate stability. The effectiveness of the model is evaluated using three datasets, i.e. Mackey-Glass, Lorenz chaotic time series of recognized benchmarks and a WIDE backbone of actual network traffic datasets. By comparing six existing model algorithms in all datasets, the results show that SAVE-AS can achieve faster convergence and high predictive accuracy while maintaining stability. Specifically, the predictive accuracy indexes such as Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) can reach a lowest optimum value of 1.1410, 0.1758 and 0.2263, and the average training time is reduced by 25.25%, 23.87% and 41.36%, respectively. The findings demonstrate that the proposed mechanism can predict network traffic more stably, accurately and rapidly in a short time regardless of time intervals or data sequence behavior. Consequently, it can provide

effective security warning guidance for network management as well as further improve network service quality.

ABSTRAK

MEKANISME PENGURAIAN MODAL RANGKAIAN NEURAL DALAM MERAMAL TRAFFIC RANGKAIAN

Ramalan trafik rangkaian adalah penting untuk pengurusan rangkaian yang berkesan kerana ia boleh memberi amaran awal kepada pentadbir sebelum insiden berlaku. Kajian ini merekabentuk satu model ramalan trafik rangkaian yang baru iaitu SAVE-AS. Ia menyertakan algoritma Scalable Artificial Bee Colony (SABC) yang baru dicadangkan, Pembinaan Ruang Fasa, Penguraian Mod Variasi (VMD) dan Mesin Pembelajaran Ekstrim Bersepadu (ELM). Mekanisma yang dicadangkan ini bermula dengan menggunakan SABC untuk mengemas kini model dengan penyelesaian baru dan menala gangguan dalam setiap iterasi untuk menangani gangguan agar dapat mencari nilai-nilai terbaik yang juga secara serentak optimum. SAVE-AS kemudian membina operator pemilihan penyesuaian. Ia secara adaptif memilih jumlah set data selepas penguraian pengoptimuman VMD untuk menetapkan dengan tepat jumlah nod lapisan tersembunyi dalam ELM untuk meningkatkan ketepatan ramalan. Sementara itu, model ELM dilatih menggunakan pelbagai jujukan data yang memenuhi keperluan untuk meminimumkan kompleksiti komputasi dalam pemodelan. Selain itu, mekanisma ini menghilangkan jujukan yang lemah disebabkan oleh kerentanan keputusan untuk mempercepatkan kadar kestabilan penumpuan. Keberkesanan model ini dinilai menggunakan tiga set data, iaitu Mackey-Glass, Lorenz Siri masa huru-hara Lorenz sebagai piawaian yang diiktiraf dan set data trafik rangkaian sebenar, WIDE. Dengan membandingkan enam algoritma model yang sedia ada dalam semua set data, hasil kajian menunjukkan bahawa SAVE-AS dapat mencapai penumpuan yang lebih cepat dan ketepatan ramalan yang tinggi sambil mengekalkan kestabilan. Secara khusus, indeks ketepatan ramalan seperti Ralat Peratus Mutlak Rata-Rata (MAPE), Ralat Mutlak Rata-Rata (MAE) dan Ralat Purata Kuasa Dua (RMSE) boleh mencapai nilai optimum terendah masing-masing sebanyak 1.1410, 0.1758 dan 0.2263, dan masa

latihan purata dikurangkan sebanyak 25.25%, 23.87% dan 41.36% masing-masing. Dapatan kajian ini menunjukkan bahawa mekanisma yang dicadangkan dapat meramalkan trafik rangkaian dengan lebih stabil, tepat dan cepat dalam masa yang singkat tanpa mengira selang masa atau kelakuan rangkaian data. Oleh itu, ia dapat memberikan panduan amaran keselamatan yang berkesan untuk pengurusan rangkaian serta meningkatkan lagi kualiti perkhidmatan rangkaian.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	-	Artificial Bee Colony algorithm
ANN	-	Artificial Neural Network
APSO	-	Adaptive Particle Swarm Optimization
AR	-	Auto Regressive
ARMA	-	Auto-Regressive Moving Average
ARIMA	-	Auto Regressive Integrated Moving Average
AS	-	Adaptive Selection
BA	-	Bat Algorithm
BP	-	Back Propagation
BPTT	-	Back Propagation Trough Time
CAIDA	-	Cooperative Association for Internet Data Analysis
CEEMD	-	Complementary Ensemble Empirical Mode Decomposition
CT	-	Chaos Theory
CVNN	-	Complex-valued Neural Network
CNN	-	Convolutional Neural Networks
DBN	-	Deep Belief Network
DE	-	Differential Evolution
DR	-	Dynamic Reservoir
DWT	-	Discrete Wavelet Transform

EEMD - Ensemble Empirical Mode Decomposition

EELM - Efficient Extreme Learning Machine

ELM - Extreme Learning Machines

EM - Error Minimized

EMD - Empirical Mode Decomposition

ESN - Echo State Network

EWT - Empirical Wavelet Transform

FA - Firefly Algorithm

FARIMA - Fractional Auto Regressive Integrated Moving Average

FOA - Fruit Fly Optimization Algorithm

FP - Fitness Prediction

FS - Free Search

GA - Genetic Algorithm

GM - Gaussian model

GMDH - Group Method of Data Handling

GPR - Gaussian Process Regression

GSA - Gravity search algorithm

GST - Grey System Theory

HHT - Hilbert-Huang Transform

IABC - Improved Artificial Bee Colony

ICA - Imperialist Competitive Algorithm

IEPM - Internet End-to-End Performance Monitoring

IFOA - Improved Fruit Optimization Algorithm

IFS - Improved Free Search

IGSA - Improved Gravity search algorithm

IHS - Improved Harmony Search

IMF - Intrinsic Mode function

IPPM - IP Performance Metrics

IPSO - Improved Particle Swarm Optimization

ITU - International Telecommunication Union

LMD - Local Mean Decomposition

LSSVM - Least Squares Support Vector Machines

LSTM - Long short-term Memory

LVCMFOA - Levy flight function and cloud generator Fruit Fly Optimization Algorithm

MA - Moving Average

MAE - Mean Absolute Error

MAD - Mean Absolute Deviation

MAPE - Mean Absolute Percentage Error

MAWI - Measurement and Analysis on the WIDE Internet

NNT - Neural Network Theory

OBL - Opposition-Based Learning

PSO - Particle Swarm Optimization

PSR	-	Phase Space Reconstruction
QFOA	-	Quantum Fruit Optimization Algorithm
QGA	-	Quantum Genetic Algorithm
QNN	-	Quantum neural network
QPSO	-	Quantum Particle Swarm Optimization
RBF	-	Radial Basis Function
RMSE	-	Root Mean Squared Error
RNN	-	Recurrent Neural Network
SABC	-	Scalable Artificial Bee Colony
SAVE-AS	-	SABC-VMD-AS-PSR-ELM
SCS	-	Spatiotemporal Compressive Sensing
SLAC	-	Stanford Linear Accelerator Center
SLFN	-	Single Hidden Layer Feedforward Network
SLT	-	Seasonal Loess Trend
ST	-	Statistical Theory
ST	-	Seasonal Transform
SVM	-	Support Vector Machines
TSD	-	Time Series Decomposition
VMD	-	Variational Mode Decomposition
WA	-	Wavelet Analysis
WD	-	Wavelet Decomposition
WT	-	Wavelet Transform

CHAPTER 1

INTRODUCTION

1.1 Overview and Motivation

With the rapid expansion of information in the 21st century coupled with increasing internet penetration rate worldwide, the Internet has consequently become a part and parcel of our everyday life. According to the International Telecommunication Union's (ITU) 2022 Annual Internet Survey report, the global Internet user-base continues to increase rapidly from 2013 to 2022. As of 29 January 2022 , the number of global Internet users reached 4.95 billion, accounting for 62.5% of the world's population.

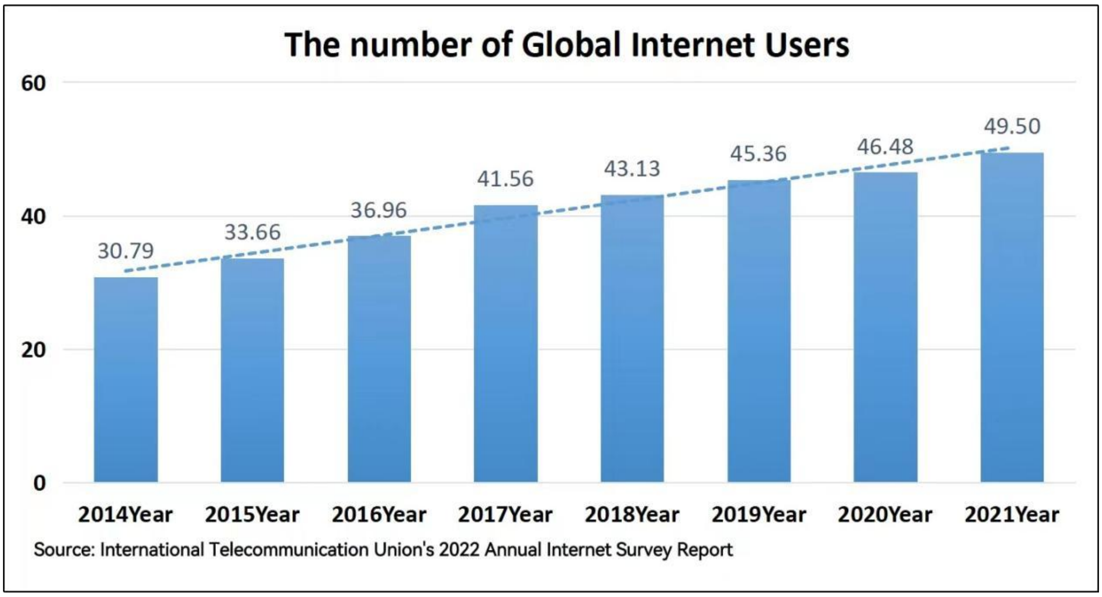

Figure 1.1 : The number of Global Internet Users