

SpringerBriefs in Computer Science

Ting Wang · Bo Li · Mingsong Chen · Shui Yu



**Machine Learning
Empowered Intelligent
Data Center Networking**
Evolution, Challenges and
Opportunities

SpringerBriefs in Computer Science

Series Editors

Stan Zdonik, Brown University, Providence, RI, USA

Shashi Shekhar, University of Minnesota, Minneapolis, MN, USA

Xindong Wu, University of Vermont, Burlington, VT, USA

Lakhmi C. Jain, University of South Australia, Adelaide, SA, Australia

David Padua, University of Illinois Urbana-Champaign, Urbana, IL, USA

Xuemin Sherman Shen, University of Waterloo, Waterloo, ON, Canada

Borko Furht, Florida Atlantic University, Boca Raton, FL, USA

V. S. Subrahmanian, University of Maryland, College Park, MD, USA

Martial Hebert, Carnegie Mellon University, Pittsburgh, PA, USA

Katsushi Ikeuchi, University of Tokyo, Tokyo, Japan

Bruno Siciliano, Università di Napoli Federico II, Napoli, Italy

Sushil Jajodia, George Mason University, Fairfax, VA, USA

Newton Lee, Institute for Education, Research and Scholarships, Los Angeles, CA, USA

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.


****Indexing: This series is indexed in Scopus, Ei-Compendex, and zbMATH ****

Ting Wang • Bo Li • Mingsong Chen • Shui Yu


Machine Learning Empowered Intelligent Data Center Networking

Evolution, Challenges and Opportunities

 Springer

Ting Wang 
Software Engineering Institute
East China Normal University
Shanghai, China

Mingsong Chen
Software Engineering Institute
East China Normal University
Shanghai, China

Bo Li 
Software Engineering Institute
East China Normal University
Shanghai, China

Shui Yu 
School of Computer Science
University of Technology Sydney
Ultimo, NSW, Australia

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-981-19-7394-9 ISBN 978-981-19-7395-6 (eBook)
<https://doi.org/10.1007/978-981-19-7395-6>

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

To support the needs of ever-growing cloud-based services, the number of servers and network devices in data centers is increasing exponentially, which in turn results in high complexities and difficulties in network optimization. To address these challenges, both academia and industry turn to artificial intelligence technology to realize network intelligence. To this end, a considerable number of novel and creative machine learning-based (ML-based) research works have been put forward in recent few years. Nevertheless, there are still enormous challenges faced by the intelligent optimization of data center networks (DCNs), especially in the scenario of online real-time dynamic processing of massive heterogeneous services and traffic data. To the best of our knowledge, there is a lack of systematic and original comprehensive investigations with in-depth analysis on intelligent DCN. To this end, in this book, we comprehensively investigate the application of machine learning to data center networking and provide a general overview and in-depth analysis of the recent works, covering flow prediction, flow classification, load balancing, resource management, energy management, routing optimization, congestion control, fault management, and network security. In order to provide a multi-dimensional and multi-perspective comparison of various solutions, we design a quality assessment criteria called REBEL-3S to impartially measure the strengths and weaknesses of these research works. Moreover, we also present unique insights into the technology evolution of the fusion of data center networks and machine learning, together with some challenges and potential future research opportunities.

Shanghai, P.R. China
March 2022

Ting Wang
Bo Li
Mingsong Chen
Shui Yu

Acknowledgments

This work was partially supported by the grants from National Key Research and Development Program of China (2018YFB2101300), Natural Science Foundation of China (61872147), the Dean's Fund of Engineering Research Center of Software/Hardware Co-design Technology and Application, Ministry of Education (East China Normal University), and the grants from Shenzhen Science and Technology Plan Project (CJGJZD20210408092400001).

Contents

1	Introduction	1
	References	5
2	Fundamentals of Machine Learning in Data Center Networks	9
2.1	Learning Paradigm	9
2.2	Data Collection and Processing	10
2.2.1	Data Collection Scenarios	11
2.2.2	Data Collection Techniques	11
2.2.3	Feature Engineering	11
2.2.4	Challenges and Insights	12
2.3	Performance Evaluation of ML-Based Solutions in DCN	13
	References	14
3	Machine Learning Empowered Intelligent Data Center	
	Networking	15
3.1	Flow Prediction	15
3.1.1	Temporal-Dependent Modeling	16
3.1.2	Spatial-Dependent Modeling	16
3.1.3	Discussion and Insights	17
3.2	Flow Classification	17
3.2.1	Supervised Learning-Based Flow Classification	22
3.2.2	Unsupervised Learning-Based Flow Classification	23
3.2.3	Deep Learning-Based Flow Classification	23
3.2.4	Reinforcement Learning-Based Flow Classification	24
3.2.5	Discussion and Insights	24
3.3	Load Balancing	28
3.3.1	Traditional Solutions	29
3.3.2	Machine Learning-Based Solutions	29
3.3.3	Discussion and Insights	30
3.4	Resource Management	31
3.4.1	Task-Oriented Resource Management	36
3.4.2	Virtual Entities-Oriented Resource Management	37

3.4.3	QoS-Oriented Resource Management	37
3.4.4	Resource Prediction-Oriented Resource Management	38
3.4.5	Resource Utilization-Oriented Resource Management	38
3.4.6	Discussion and Insights	39
3.5	Energy Management	46
3.5.1	Server Level	47
3.5.2	Network Level	47
3.5.3	Data Center Level	48
3.5.4	Discussion and Insights	49
3.6	Routing Optimization	54
3.6.1	Intra-DC Routing Optimization	55
3.6.2	Inter-DC Routing Optimization	55
3.6.3	Discussion and Insights	55
3.7	Congestion Control	61
3.7.1	Centralized Congestion Control	62
3.7.2	Distributed Congestion Control	63
3.7.3	Discussion and Insights	63
3.8	Fault Management	64
3.8.1	Fault Prediction	69
3.8.2	Fault Detection	70
3.8.3	Fault Location	70
3.8.4	Fault Self-Healing	70
3.8.5	Discussion and Insights	71
3.9	Network Security	75
3.10	New Intelligent Networking Concepts	79
3.10.1	Intent-Driven Network	80
3.10.2	Knowledge-Defined Network	80
3.10.3	Self-Driving Network	81
3.10.4	Intent-Based Network (Gartner)	82
3.10.5	Intent-Based Network (Cisco)	83
	References	84
4	Insights, Challenges and Opportunities	101
4.1	Industry Standards	105
4.1.1	Network Intelligence Quantification Standards	105
4.1.2	Data Quality Assessment Standards	105
4.2	Model Design	105
4.2.1	Intelligent Resource Allocation Mechanism	105
4.2.2	Inter-DC Intelligent Collaborative Optimization Mechanism	106
4.2.3	Adaptive Feature Engineering	106
4.2.4	Intelligent Model Selection Mechanism	106
4.3	Network Transmission	106
4.4	Network Visualization	107
	References	108

Contents	xi
5 Conclusion	109
Index	111

Acronyms

AIMD	Additive Increase Multiplicative Decrease
ANN	Artificial Neural Network
AR	Augmented Reality
ARIMA	Autoregressive Integrated Moving Average Model
ARMA	Autoregressive Moving Average Model
BGP	Border Gateway Protocol
BI	Blocking Island
BMP	BGP Monitoring Protocol
Bof	Bag of Flow
CC	Congestion Control
CFD	Computational Fluid Dynamics
CNN	Convolutional Neural Networks
CRAC	Computer Room Air Conditioner
CRE	Cognitive Routing Engine
DBN	Deep Belief Network
DC	Data Center
DCN	Data Center Network
DDoS	Distributed Denial of Service Attack
DDPG	Deep Deterministic Policy Gradient
DL	Deep Learning
DNN	Deep Neural Network
DPI	Deep Packet Inspection
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
DT	Decision Tree
ELM	Extreme Learning Machine
eMBB	Enhanced Mobile Broadband
eMDI	Enhanced Media Delivery Index
eMTC	Enhanced Machine-Type Communication
FCT	Flow Completion Time
FPGA	Field Programmable Gate Array

FTR	Fundamental Theory Research
GBDT	Gradient Boosting Decision Tree
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
GWO	Grey Wolf Optimization
HSMM	Hidden Semi-markov Model
IANA	Internet Assigned Numbers Authority
IDS	Intrusion Detection System
IT	Information Technology
LA	Automatic Learning
LSTM	Long Short-Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ME	Mean Error
ML	Machine Learning
MSE	Mean Squared Error
NBD	Naïve Bayes discretization
NFV	Network Functions Virtualization
NIDS	Network Intrusion Detection System
NMSE	Normalized Mean Square Error
NN	Neural Network
O&M	Operations and Maintenance
PAC	Packaged Air Conditioner
PC	Personal Computer
QoS	Quality of Service
RAE	Relative Absolute Error
RF	Random Forest
RFR	Random Forest Regression
RL	Reinforcement Learning
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RRMSE	Relative Root Mean Squared Error
RRSE	Relative Root Squared Error
RSNE	Ratio of Saved Number of Entries
RTT	Round-Trip Time
SDN	Software Defined Network
SL	Supervised Learning
SLA	Service-Level Agreement
SNMP	Simple Network Management Protocol
SVM	Support Vector Machine
SVR	Support Vector Regression
TCP	Transmission Control Protocol
TPU	Tensor Processing Unit
TWAMP	Two-Way Active Measurement Protocol