SpringerBriefs in Computer Science

Ting Wang · Bo Li · Mingsong Chen · Shui Yu

# Machine Learning Empowered Intelligent Data Center Networking

## Evolution, Challenges and Opportunities

Springer

# SpringerBriefs in Computer Science

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

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

Mingsong Chen
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



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

# 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 |