



Shaoyuan Huang

Ph.D. 4th year, College of Intelligence and Computing, Tianjin University, Tianjin
[dblp](#) [Google Scholar](#) [homepage](#)

Building No. 55, Tianjin University, Haihe Education District, Jinnan, Tianjin
hsy_23@tju.edu.cn +86-15022618263

Research Field

- Distributed System Workload and Performance Modeling
- AI Inference Serving Systems
- Resource Provisioning

Education Experience

- Visting Ph.D.(2024-2025) Department of Engineering, King's College London
(Supervisor: Prof. Yansha Deng)
- Ph.D. (2022-Now), M.S.(2020-2022), B.S. (2016-2020)
From College of Intelligence and Computing, Tianjin University, Tianjin, China
(Supervisor: Prof. Xiaofei Wang, Peiyang Young Scholar, National Thousand Youth Talents Plan)

Internship

- 2021.09-2022.06 Algorithm Development Intern, in PPIO Cloud Computing (Shanghai) Co.
 - Designed the workload and utilization prediction model based on Xgboosting and residual learning, with an accuracy of over 90% across thousands of servers.
 - Participated in the development of a prototype prediction-based task deployment recommendation system, responsible for algorithm integration and data flow.

Selected Publications

- **Journal**
 1. Shaoyuan Huang, Zheng Wang, Heng Zhang, Xiaofei Wang, Cheng Zhang, Wenyu Wang "DynEformer: A Unified Framework for Robust Workload Prediction Under Dynamic Environment," in *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2024. Under Review.
 2. Shaoyuan Huang, Heng Zhang, Xiaofei Wang*, Min Chen, Jianxin Li, Victor C.M. Leung "Fine-grained Spatio-Temporal Distribution Prediction of Mobile Content Delivery in 5G Ultra-Dense Networks," in *IEEE Transactions on Mobile Computing (TMC)*, 2022. (JCR-1, IF:7.9)
 3. Shaoyuan Huang, Yuxi Zhang, Guozheng Peng, Juan Zhao, Keping Zhu, Heng Zhang, Xiaofei Wang*, "MF-GCN-LSTM: A Cloud-Edge Distributed Framework for Key Positions Prediction in Grid Projects," in *Journal of Cloud Computing*, 2022. (JCR-2, IF:4.0)
 4. Heng Zhang, Shaoyuan Huang, Xin Wang, Jianxin Li, Xiaofei Wang*, Victor C. M. Leung, "A Measurement-driven Analysis and Prediction of Content Propagation in the Device-to-Device Social Networks," in *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2022. (JCR-1, IF:8.9)
- **Conference**
 1. Shaoyuan Huang, Tiancheng Zhang, Zhongtian Zhang, Xiaofei Wang, Lanjun Wang, Xin Wang, "MetaEformer: Unveiling and Leveraging Meta-Patterns for Complex and Dynamic Systems Load Forecasting", in *31TH ACM SIGKDD*

- Conference on Knowledge Discovery and Data Mining (ACM SIGKDD)*, 2025, (CCF-A).
2. **Shaoyuan Huang**, Zheng Wang, Zhongtian Zhang and Heng Zhang, Xiaofei Wang, Wenyu Wang, "[Seer: Proactive Revenue-Aware Scheduling for Live Streaming Services in Crowdsourced Cloud-Edge Platforms](#)", in *IEEE International Conference on Computer Communications (IEEE INFOCOM)*, 2024, (CCF-A).
 3. **Shaoyuan Huang**, Zheng Wang, Heng Zhang, Xiaofei Wang, Cheng Zhang and Wenyu Wang, "[One for All: Unified Workload Prediction for Dynamic Multi-tenant Edge Cloud Platforms](#)", in *29TH ACM SIGKDD Conference on Knowledge Discovery and Data Mining (ACM SIGKDD)*, 2023, (CCF-A).
 4. **Shaoyuan Huang**, Heng Zhang, Xiaofei Wang, Min Chen, Jianxin Li, Victor C.M. Leung, "[Spatial-Temporal-Social Multi-Feature-based Fine Grained Hot Spots Prediction for Content Delivery Services in 5G Era](#)", in *30th ACM International Conference on Information and Knowledge Management (ACM CIKM)*, 2021, (CCF-B).
 5. Yuting Li, **Shaoyuan Huang**, Tengwen Zhang Cheng Zhang Xiaofei Wang and Victor C.M. Leung, "[Sentinel: Scheduling Live Streams with Proactive Anomaly Detection in Crowdsourced Cloud-Edge Platforms](#)", in *IEEE International Conference on Computer Communications (IEEE INFOCOM)*, 2025, (CCF-A).
 6. Heng Zhang, **Shaoyuan Huang**, Mengwei Xu, Deke Guo, Xiaofei Wang, Victor C. M. Leung and Wenyu Wang, "[How Far Have Edge Clouds Gone? A Spatial-Temporal Analysis of Edge Network Latency In the Wild](#)", in *IEEE/ACM International Symposium on Quality of Service (IWQoS)*, 2023, (CCF-B).
 7. Tiancheng Zhang, **Shaoyuan Huang**, Cheng Zhang, Xiaofei Wang, Wenyu Wang, "[EasyTS: The Express Lane to Long Time Series Forecasting](#)", in *AAAI 2024 Demonstration Program*, 2024, (CCF-A).
- **Latest Work:**
 1. **Huang et al.**, "[CLIF: Co-Orchestrating for Joint Large Language Model Inference Serving and Fine-tuning in GPU Clusters](#)" (*NSDI2026*, Under Review)
 2. **Huang et al.**, "[CoLLM: A Unified Framework for Co-execution of LLMs Federated Fine-tuning and Inference](#)" (*INFOCOM2026*, Under Review)

Talk

- 31th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (ACM SIGKDD), 2025, Toronto, Canada, participation and presentation.
- 44th IEEE International Conference on Computer Communications (IEEE INFOCOM) 2025, London, United Kingdom, participation and presentation.
- 43th IEEE International Conference on Computer Communications (IEEE INFOCOM) 2024, Vancouver, Canada, participation and presentation.
- 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (ACM SIGKDD), 2023, Long Beach, CA, USA, online participation and presentation.
- IEEE Global Communications Conference (Globecom), 2023, Kuala Lumpur, Malaysia, participation and presentation.

- 30th ACM International Conference on Information and Knowledge Management (ACM CIKM), 2021, online participation and presentation.

Academic Service

- **Session Chair** on Workshop on Integrating Edge Intelligence and Large Model in Next Generation Networks (IEILM'24, Colocated with INFOCOM'24)
- **Organizing Volunteer**, IEEE International Conference on Computer Communications (INFOCOM'25)
- **Reviewer**
 - SIGKDD 2023, 2024, 2025
 - IEEE Conference on Vehicular Technology (VTC)
 - IEEE Transactions on Mobile Computing (TMC)
 - IEEE Transactions on Machine Learning in Communications and Networking
 - IEEE Network Magazine
 - Artificial Intelligence Review

Patent

- "Multi-feature based neural network for content delivery hotspots prediction", Chinese Patent, CN112822045B (Patent Authorized)
- "Edge cloud server utilization prediction method, prediction device and storage medium based on boosting algorithm", Chinese Patent, CN114721898A (Patent Authorized)

Including those not listed, totaling 12 patents.

Award

- SIGKDD2025 Student Travel Grant
- 2025, BYD Scholarship (only 7 PhDs in shcool)
- 2024, China Scholarship Council
- 2024, CCF DPCS Distinguished Doctorate (nationwide: 6 people)
- 2021, 2023, "Suzhou Talent Scholarship", Suzhou Government Talent Group (20 students in shcool , 1 student in the college)
- 2020, "Outstanding Graduate" of Tianjin University
- 2017-2019, 2023 "Merit Student" of Tianjin University

Technical Contribution

- **Open source system models and datasets**
 - [ECAI25 HRS: Hybrid Representation Time Series Forecasting](#)
 - [KDD25 MetaEformer: MetaEformer v1.0 - Initial Release](#)
 - [Edge Cloud Server Latency Measurements](#)
 - [DynEformer: Edge Cloud Server Workload Prediction Framework](#)
 - [ECW: Edge Cloud Server Workload Dataset](#)