Jinzhen Wang

PhD Candidate in Computer Engineering
Department of Electrical and Computer Engineering
New Jersey Institute of Technology

Email: jw447@njit.edu
Mobile: (551)580-4404
Website: https://jw447.github.io

RESEARCH INTERESTS

High Performance Computing, Data Compression, Scientific Data Analysis, Distributed Computing

EDUCATION

New Jersey Institute of Technology Newark, New Jersey, US
Jan 2018 - Now Ph.D. Computer Engineering (Dissertation advisor: Dr. Qing Liu)

New Jersey Institute of Technology Newark, New Jersey, US
Sept 2015 - May 2017 M.S. Electrical Engineering

Shandong University Jinan, Shandong, China
Sept 2011 - Jul 2015 B.S. Internet of Things (Electrical Engineering)

RESEARCH EXPERIENCE

High-Ratio Lossy Compression: Exploring the Autoencoder to Compress Scientific Data (IEEE TBD)

Advisor: Dr. Qing Liu

Motivation:

Auto-encoder has been widely used in image compression and dimension reduction and achieves superior performance, motivating the adoption in the use of scientific data compression.

Contribution:

1. Propose and design the first error-bounded auto-encoder compression framework that achieves the superior reduction performance than traditional scientific lossy compression techniques.
2. Optimize the auto-encoder compression framework via various data pre-processing techniques (e.g., normalization) and combining with backend lossless encoding techniques.
3. Evaluate the compression framework over 18 datasets of various scientific simulations from different disciplines to show the generality.

Estimating Lossy Compressibility of Scientific Data Using Deep Neural Networks (IEEE LOC)

Advisor: Dr. Qing Liu

Motivation:

The level of compression to be expected is usually hard to know, making it hard for domain scientists to make informed decisions on the tradeoff between data accuracy and compression performance.

Contribution:

1. Understand and characterize the HPC compression performance using deep learning.
2. Propose and design a deep-learning-based framework that estimate the compressibility of HPC scientific data. Demonstrate that compressor-specific features drastically improve the prediction accuracy.
3. Evaluate the proposed work with sampling-based and analytical model that were used to estimate the compression performance.

Compression Ratio Modeling and Estimation Across Error Bounds for Lossy Compression (IEEE TPDS)

Advisor: Dr. Qing Liu

Motivation:

The lossy compression techniques used in HPC community suffer from suboptimal performance, since users lack the guidance to configure error bound values.

Contribution:

1. Study and demonstrate the relation between lossy compression error bounds and compression ratios. Develop insights on the interplay between compression inner metrics and error bound.
2. Propose an analytic model to predict the compression ratio for a lossy error bound based on compression metrics.
3. Evaluation using 20 real-world scientific datasets on two leading lossy compression techniques (SZ and ZFP).

PUBLICATIONS

- Liu, T., **Wang, J.**, Liu, Q., Alibhai, S., Lu, T., & He, X. (2021). High-ratio lossy compression: Exploring the autoencoder to compress scientific data. IEEE Transactions on Big Data.
- Qin, Z., **Wang, J.**, Liu, Q., Chen, J., Pugmire, D., Podhorszki, N., & Klasky, S. (2020). Estimating Lossy Compressibility of Scientific Data Using Deep Neural Networks. IEEE Letters of the Computer Society, 3(1), 5-8.
- **Wang, J**., Liu, T., Liu, Q., He, X., Luo, H., & He, W. (2019). Compression ratio modeling and estimation across error bounds for lossy compression. IEEE Transactions on Parallel and Distributed Systems, 31(7), 1621-1635.
- Liu, T., Alibhai, S., **Wang, J.**, Liu, Q., He, X., & Wu, C. (2019, August). Exploring transfer learning to reduce training overhead of hpc data in machine learning. In 2019 IEEE International Conference on Networking, Architecture and Storage (NAS) (pp. 1-7). IEEE.
- Qiao, Z., Lu, T., Luo, H., Liu, Q., Klasky, S., Podhorszki, N., & **Wang, J.** (2018). Sirius: Enabling progressive data exploration for extreme-scale scientific data. IEEE Transactions on Multi-Scale Computing Systems, 4(4), 900-913.

SKILLS

Programming language: Python, C, C++

High Performance Computing platform: Summit@ORNL, Cori@NERSC.

Cloud Computing Platform: Google Cloud Platform, AWS

Software framework: Tensor-Flow, Pytorch, Scikit-learn