

Domain-Specific Compression Using Auto-Encoders For Climate Data

- ▶ Author: Ravi Mallikarjun Yadav Chennaboina
- ▶ Type: Master's Thesis
- ▶ Date: 2022-05-18
- ▶ Reviewers: Jun.-Prof. Dr. Michael Kuhn, Dr. Jakob Lüttgau
- ▶ Supervisors: Jun.-Prof. Dr. Michael Kuhn, Dr. Jakob Lüttgau
- ▶ Download: PDF

The importance of climate data is spread across different sectors like weather forecasting, agriculture, water resource management etc. Moreover, climate data collected over the years helps to analyse the change in climate patterns and provides crucial information about the future climate. However, collecting such data sets up a massive challenge for the limited storage facilities. Compressing data is one of the solutions to the storage problem, which reduces the size of data by removing redundant information. This work aims to understand the challenges behind different Autoencoder architectures achieving a high compression ratio while maintaining the originality of data when reconstructing the climate data. The employed climate dataset is from the open-source Weatherbench dataset containing 14 climate variables. Evaluation metrics like compression ratio, Structural Similarity Metrics and Peak Signal to Noise Ratio are used to measure and select the best performing architecture on the climate variables. The Autoencoders show good reconstruction results for the variables geopotential, potential vorticity, vorticity, total incident solar radiation, temperature and 2m temperature and relative humidity, 10m u component of wind, 10m v component of wind, u component of wind and v component of wind but worse on variable total cloud cover. Variational Autoencoders achieve the highest compression ratio of 43.29:1 and better reconstruction quality compared to other Autoencoder architectures. Variational Autoencoder compresses 14 times more than that of other lossless techniques like SZ, ZFP and PCA. The compression and decompression speed of lossless compression techniques like Zstd, Zlib, and Lz4 turns out to be 10-17 times faster than the Variational Autoencoder.