ABSTRACT

The challenges of storing, retrieval and analysis of big weather datasets from cloud storage poses a great barrier to researchers and decision makers. Efficient data compression techniques can help in mitigating storage constraints. One approach is to maintain a precomputed summary in a local machine and generate enough data using prediction models, which allows users to visualize data in a small machine, even when the user is offline. To achieve this goal, we first designed features and corresponding machine learning models, and then analyzed how well the models can approximate the visualization compared to the visualization computed from the real dataset (WRF model output).

We built two models: The first model learns from ALBEDO, EMISS, GRDFLX features over a geographical area and predicts the SOIL MOISTURE. The second model learns from four features, where we assume that some SOIL MOISTURE data is missing. Thus the first (second) model works with all (resp., 50%) SOIL MOISTURE data missing, and reduces the storage space requirement by 25% (resp., 12.5%). The predicated visualization was reasonable approximation to the original.

MODEL ARCHITECTURE

Each feature vector at a latitude and longitude pair containing 3 variables: ALBEDO, EMISS and GRDFLX along with 8 of their neighbors: total 9 values each. NO Soil MOISTURE value was given while training the model. The training and testing datasets had data of 15 and 5 days of the year 2014 respectively. The objective was to predict SOIL MOISTURE of all the points of the dataset.

MODEL 1: DATA PREDICTION

Figure 1: Original Data in the left column (a, c) with the predicted ones in the right (b, d)

MODEL 2: DATA COMPRESSION

Only odd rows and columns were taken as feature vectors, even ones were discarded as the step of data compression.

Figure 1: Original Data in the left column (a, c) with the predicted ones in the right (b, d)

REFERENCES
