April 2022

FREQUENCY ANALYSIS TO FACILITATE THE LIGHTWEIGHT TRANSFER OF TIMESERIES DATA

Thomas Vegas

Anirban Karmakar

FREQUENCY ANALYSIS TO FACILITATE THE LIGHTWEIGHT TRANSFER OF TIMESERIES DATA

AUTHORS:
Thomas Vegas
Anirban Karmakar

ABSTRACT

Internet of Things (IoT) devices can collect diverse amount of data, which can be represented as timeseries data. Such data sent can be aggregated and again sent by intermediate parties. Often the properties of a data set are related to physical phenomena. It would be useful to take advantage of similarities and signal aggregations in order to reduce the size of data transfers. Presented herein are techniques through which multiple data measurements can be clustered together as a two-dimensional matrix. By clustering data measurements in such a manner, simple image processing techniques can be applied to the data such that the data can be transferred under an improved compressed form.

DETAILED DESCRIPTION

When considering a set of timeseries data, the data may be represented in a manner such that the amount of data can be compressed using a high compression ratio without making many assumptions regarding the format of the data.

Consider, for example, a set of timeseries data that may be reflective of many temperature measurements across many cities, over the same period of time. Such data may be represented as a two-dimensional (2D) matrix in which the X-axis can indicate time, while the Y-axis can indicate a measurement considered (e.g., a City for temperature measurements).

Presented herein are techniques through which a frequency domain representation may be provided for such timeseries data. For example, to obtain a more localized frequency coefficient, the various measurements can be sorted using the following approach:

1. Take first measurement timeseries, use it as current measurement;
2. Output a current measurement;

3. Find, among the remaining timeseries data, a data point that has the highest correlation with respect to the current measurement; and
4. Use the located data point as the current measurement, and repeat the process beginning from Step 2 until there are no more measurements left in the timeseries data.

Next, frequency processing can be performed through batching, which provides to the removal of high frequency coefficients. Using that representation, data can be decomposed in frequency coefficients using a standard 2D discrete cosine transform (DCT). The data can then be truncated to zero high frequency coefficients that are representing noise on the time domain data, and each of the coefficients can be quantized within an 8-bit range. Thus, the data that can be transmitted is actually the frequency coefficients, such as a coefficient matrix sent diagonally, to localize similar values in which the coefficients are sent as 8-bits values. On the receiving side, a receiver can perform the inverse processing in order to recover the data in the time domain.

As compared to traditional Fast Fourier Transform (FFT) processing, such processing is generally applied to one-dimensional data (e.g., evolution of temperature over a period of time. In contrast, DCT processing is known to be able to be applied to two-dimensional datasets, and is real/sinus based. Thus, this proposal provides for the ability to preprocess multiple sets of one-dimensional data to obtain two-dimensional data sets that will lend themselves well to DCT compression.

With regard to implications regarding the potential loss in fidelity of data due to the transformation prescribed herein, consider various error measurement graphs based on an example simulation, as shown and discussed below. Generally, in cases involving long data sets, error measurements will be minimal.

Consider, for example a simulation involving approximately 50 temperatures measurements for 19 cities and simulated using the process described above. The data represented from such processing using the simulated data set allows a 3x size reduction with trivial encoding.

Without the processing technique prescribed herein the size of the simulated data is 2128 bytes. With a simple 8-bit quantization of the simulated data (not processed using the techniques prescribed herein), the data can be reduced to 1064 bytes, which can be compressed to 988 bytes (using XZ compression).

In contrast, the simulated data, when processed in accordance with the DCT processing technique prescribed herein can be reduced to 1064 bytes, and can then be further compressed (using XZ compression) to 356 bytes.

Figure 1A, below, provides a frequency representation, which illustrates that frequency coefficients are much more similar and prone to compression. The error observed for the DCT processing on the simulated data, without any optimizations, is illustrated in Figure 1B.

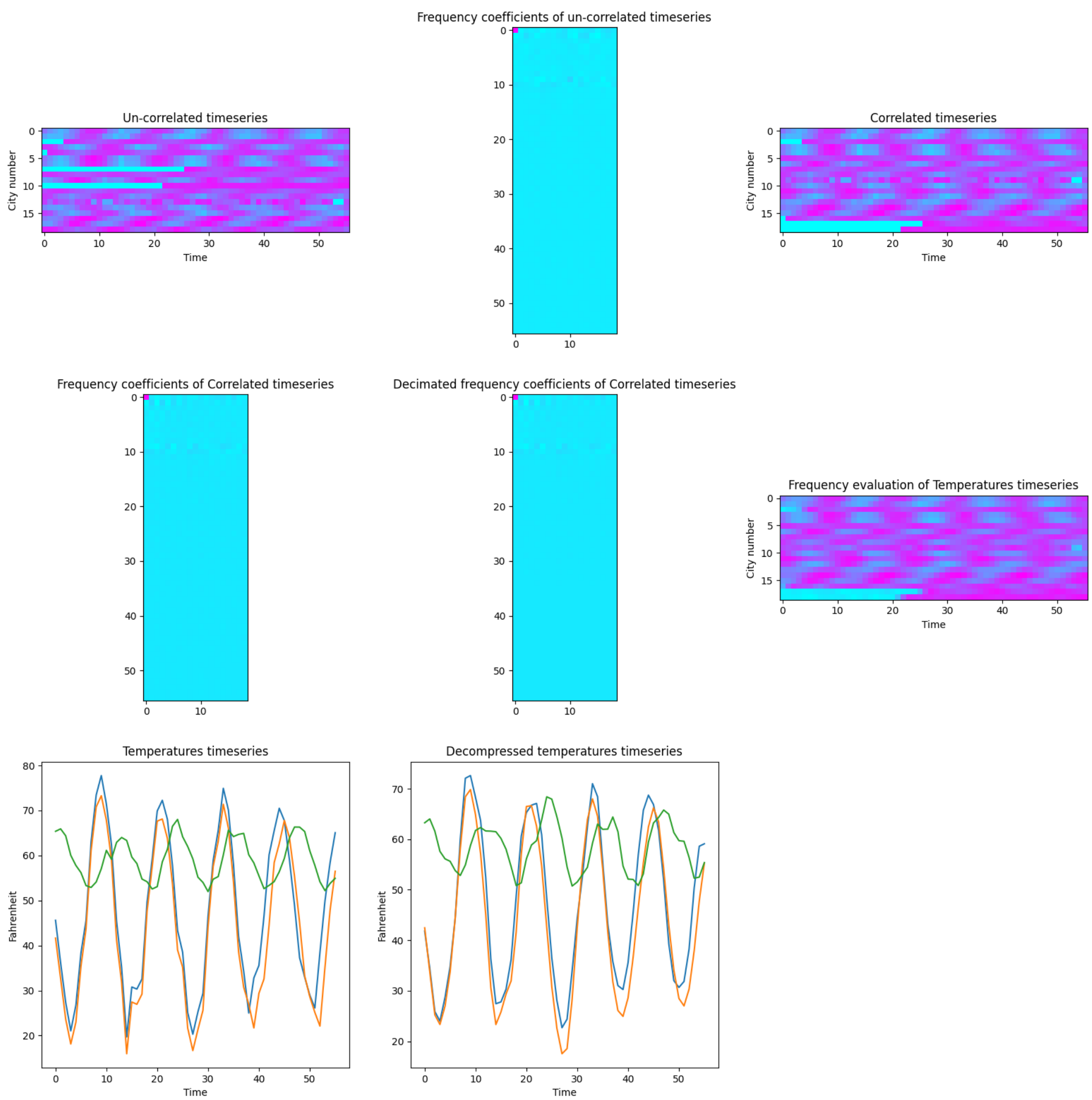

Figure 1A: Example Frequency Representation of Simulated Data

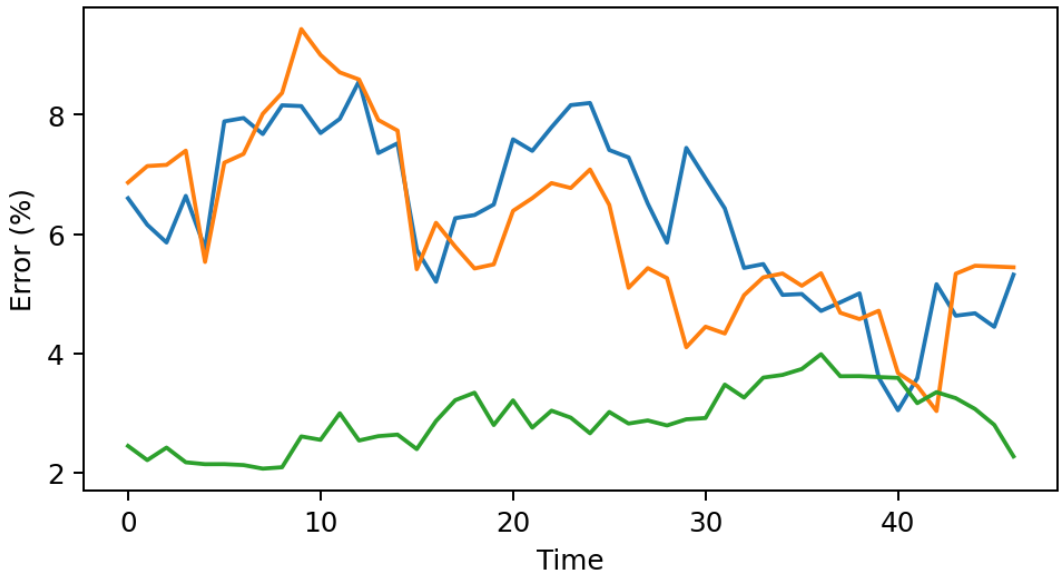

Figure 1B: Error Observed for the Simulated Data, Based on a Mean over 10 Samples

Accordingly, techniques are presented herein through which multiple data measurements can be clustered together as a two-dimensional matrix and processed using simple image processing techniques in order to reduce the size of collected data, such that it can be transferred under an improved compressed form.