

Draft: Automating Output Evaluation

Jonas Mærsk Thye Kjellerup

October 10, 2022

1 Self-similarity Matrices

A similarity matrix is used to analyse the similarity of two time series that consist of similar features. The matrix is described in terms of the similarity measure of its features $s : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}$, where \mathcal{F} is a feature space. In turn the similarity matrix $S \in \mathbb{R}^{N \times M}$ is defined as follows, where x and y are time series:

$$S(n, m) = s(x_n, y_m)$$

A self-similarity matrix (hereon abbreviated as SSM) is a similarity matrix, where both timeseries x and y are the same. Looking at the self-similarity data can be useful when trying to detect patterns and repetition within data.

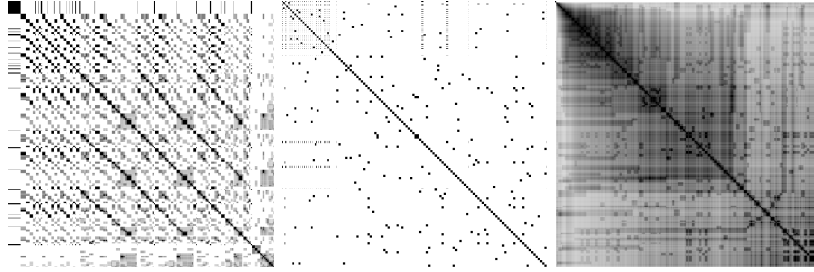


Figure 1: Visualisations of SSMs of three different datasets. The leftmost one being a MIDI file from the maestro dataset [maestro], and two other two being MIDI files output by a neural network.

In figure 1 it can be seen that the matrix is symmetric across the top-left/bottom-right diagonal. This is a result of the matrix being composed of two identical data series. This can be exploited when trying to store the digital representation of the matrix in a computer program by omitting either side of the matrix. This results in the matrix taking up memory equivalent to $N \cdot (N+1) \cdot \frac{1}{2}$ instead of N^2 .

2 Jaccard Index

The Jaccard index is a measure of similarity between two sets. It is defined as the ratio between the intersection of two sets and the union of those sets. [jaccard]

$$J(A, B) = \frac{A \cap B}{A \cup B}$$

In the context of computer programming, it can be worth noting that a series of bits can be used to represent a finite set, where each bit represents whether a given element is present in the set. This allows for the substitution of the union and intersection operations in favor of the bitwise or and bitwise and operations, which are generally easier for a computer to compute.