Contents

1 Shape of Time Series

1.1 Topic

Time Series Search Engine - With this thesis different methods are explored that can be utilized to create a time series search engine with faster retrieval times than using a DTW algorithm to compare the time series to each other.

1.1.1 First approach

- transform all time series to Fourier space via FFT

- identify the top 3-5 most important frequencies based on power spectrum

- build a band of frequencies that increases in range on a log scale - > lower frequencies are more important to be the same than higher frequencies

- order frequencies based on their level of energy

- compute proximity of series by comparing the overlap of common frequencies if frequencies overlap on the same position

- compare proximity for hourly series based on DTW total cost function

1.1.2 Issues so far

- parallelization to allow reasonable run time for comparing each series with each other series in M4 (100k series)

- spectral leakage (some frequencies seem to leak and therefore the same frequency range appears twice in the power spectrum) –> possibly apply some window-function (e.g. Hamming window)

- Does same (or just similar frequency mean that the series looks similar)?

- Does it matter whether the energy level is significantly different or not?

- How to ensure that a single frequency really describes a time series? E.g. measure variance after reconstruction? What is a good threshold for confirming whether the frequency describes the series well?

- Does time series decomposition improve the results? If so, how to integrate trends?

1.1.3 Next activities

- check with fft retransformation how much variance is removed to indicate whether this is meaningful approximation?

- compute comparison for each series and compute distance for top 5 results per series

- review Singular Spectral Decomposition as alternative to FFT

1.2 Activities so far

1.2.1 create DTW computation to find closest time series

- done for hourly M4 data set only

- 414 time series ~ 15 hours of computation

- not feasible for remaining series / not feasible for

1.2.2 compute FFT for all time series in M4

- compute FFT

- create power spectrum

- create log scaled frequency ranges to apply to the frequencies of each series

- compute 4 most powerful frequencies of each series

- compare frequency ranges for each series in M4

1.2.3 Parallelization

- find multiprocessing that allows for reasonable execution time for M4 dataset

1.3 geometric comparison (allow for different lenght)

Can I forecast a similar shape with
    What makes a similarity metric good or bad?

1.4   Similarity metrics

- serach engine for time series

- top 10 similar features

1.5   TODO Literature

- time series similarity

- time series retrieval

- clustering time series

1.6   Similarity

[?] –> important paper (check references) traditional similarity-based methods consider local geometric properties of raw time series or global topological properties propose framework which considers local geometric features and global topological characteristics of time series data

[?] -> not really relevant -> biological information

[?] -> Pixelwise studies -> not sure what they do!

[?] -> similarity measure TAOT based on optimal transport; interesting method to be tested

- currently the time adaptivity is achieved distance between time steps -> Are there alternatives?

  - method is confirmed via k-mediods check for effectiveness? How does that work?

[?] –> mixes geometric and topological properties to compare series

1.7   Measures of Similarity

1.7.1   geometric

1.7.2   topological

1.8   Idea pool

- time series analysis

1.9 Search Engine for time series

- take metrics from ts **geometric** and **topological** and compute a similarity metric

1.9.1 **TODO** measure performance cost of calculation

1.9.2 **TODO** Apply to real-world scenario -> e.g. M-competition set

1.9.3 Question: What can you do? Maybe take series with similar values (threshold?) and train model and look at forecast results?

1.9.4 **TODO** checkout UCR classification dataset

1.10 Question: How to prove that those series are actually similar? What is the control?

1.10.1 Create synthetic data sets to compute how the metrics are different when one property is changed? E.g. what is the difference b/w sine and cosine?

1.10.2 What if the amplitude is different? What if the periodicity is different?

1.11 Math concepts

1.11.1 Phase Space

- every parameter or degree of freedom is represented as an axis of multidimensional space

- one-dimensional system is called phase line, two-dimensional system is called phase plane

1. Core Idea For every possible state of the system or allowed combination of the values of the system's parameters, a point is included in the multidimensional space. The system is observed over time and the corresponding values for each dimension are traced as line (**phase space trajectory**) for this system. This can be a high-dimensional space.

   The phase space trajectory represents the set of states compatible with starting from one particular initial condition (wiki)

- for TS this means to check from different starting points (maybe cross validation)

2. Consequences

   - phase space as a whole represents all that a system can be and from the shape qualities of the system can be identified