

Time Series Clustering using Coherence

Jianwei Yue¹, Glen Takahara¹, Brian Franczak², Wesley S. Burr³

¹Department of Mathematics & Statistics, Queen's University
99 University Avenue, Kingston, Ontario, Canada
skye.griffith@queensu.ca; takahara@queensu.ca,

²Department of Mathematics & Statistics, MacEwan University,
PO Box 1796, Edmonton, Alberta
franczakb@macewan.ca

³Department of Mathematics, Trent University
1600 West Bank Drive, Peterborough, Ontario, Canada
wesleyburr@trentu.ca

Extended Abstract

Large-scale time series data sets are often interesting for both specific series reasons, and as aggregates. When considering such data as a whole, *clustering* is a convenient way to consider classification aspects of their characteristics. In this work, we describe a novel approach to time series clustering which uses a derived distance metric based on the spectral densities of the individual time series.

The spectral density of a time series is a sufficient statistic, and if estimated robustly, carries a tremendous amount of information on the structural characteristics. One commonly used robust estimation procedure is the multitaper estimate [1], which uses a family of orthogonal window (or *taper*) functions to give K approximately independent estimates, which can then be combined to give a consistent, asymptotically unbiased estimate of the spectral density. One large advantage of this approach is that it also provides access to a number of additional methods which can be built on the K estimates, including the *Magnitude-Squared Coherence* (MSC).

The MSC is a function of frequency, giving a standardized $[0, 1]$ measure of the degree of frequency-specific correlation between two time series. With application of a transformation, the MSC coefficients are approximately normal [2], and thus simple hypothesis tests can be applied to the coefficients. Of course, the usual considerations of false detection rates and multiple comparisons apply, but with careful consideration and choice of α , the method is quite reasonable for detection of bands of frequencies where two series are coherent.

Now, for clustering of time series, the MSC provides a frequency-specific estimate of similarity. By considering only statistically significant MSC coefficients (with user's choice of α), a distance metric can be constructed as $\sum_{\{j \in I\}} T(m_j)$ with I the set of significant coefficients at level α , T the transformation of the MSC to approximate normality, and m_j the j^{th} MSC, by frequency.

In this work we discuss the results of applying this distance metric to the problem of time series clustering, in a k -means framework. We explore a large-scale simulation study of the procedure against several other approaches, including classic k -means and spectral methods [3]. We show that our novel algorithm has considerably better performance than the other methods when the commonality between cluster members is expressed as coherence.

References

- [1] David J. Thomson. "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055-1096, 1982
- [2] D. J. Thomson, "Jackknifing multitaper spectrum estimates," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 20-30, 2007.
- [3] C. Shaw and G. King, "Using cluster analysis to classify time series," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1, pp. 288 - 298, 1992.