

Un método cluster de series temporales basado en frecuencias y magnitudes de picos espectrales

Borja Lafuente Rego¹ y José A. Vilar¹

¹Departamento de Matemáticas, Universidade da Coruña.

RESUMEN

Se propone un procedimiento de análisis cluster de series de tiempo en dos etapas basado en comparar frecuencias y magnitudes de los máximos absolutos de las densidades espectrales.

En la primera etapa se evalúa la disimilaridad entre cada par de series en términos del p -valor asociado a un test de igualdad de las frecuencias donde se alcanza el máximo de sus espectros. Se considera un test bootstrap propuesto por Timmer *et al.* (1999) y que trabaja como sigue. El estadístico del test es la diferencia entre las frecuencias asociadas a los máximos de las estimaciones no paramétricas de ambos espectros. Se generan nuevas realizaciones de los periodogramas basadas en la estimación espectral inicial y la distribución asintótica χ^2 de las ordenadas del periodograma, permitiendo así reestimar los espectros y obtener réplicas bootstrap del estadístico del test. Tomando como punto de partida la matriz de p -valores, se aplica el algoritmo clustering propuesto por Maharaj (2000) que, por construcción, conduce a una solución cluster donde el número de clusters es automáticamente determinado para un nivel de significación prefijado. La principal característica de este algoritmo es que dos series cualesquiera en el mismo cluster no difieren significativamente en las frecuencias donde se maximizan sus densidades espectrales.

Cada cluster de la solución obtenida en esta primera etapa, se somete entonces a un nuevo proceso clustering basado aquí en chequear si difieren las áreas bajo las densidades espectrales en un entorno local de las frecuencias pico. Como en la primera etapa, se desarrolla un test bootstrap a tales efectos y se obtienen los correspondientes p -valores que permiten aplicar el algoritmo clustering de Maharaj (2000).

La metodología clustering descrita se aplica en varios escenarios simulados, considerando diferentes modelos de series (incluyendo series lineales y series no lineales). Se observa un buen comportamiento del método, resultando claramente competitivo con otros procedimientos considerados en la literatura de clustering de series de tiempo (see Pérttega y Vilar, 2010).

Palabras y frases clave: Análisis cluster, bootstrap, análisis espectral, procesos no lineales.

REFERENCIAS

- Maharaj, E.A. (2000) Clusters of time series. *Journal of Classification* **17**: 297–314.
- Pérttega, S. y Vilar, J.A. (2010) Comparing Several Parametric and Nonparametric Approaches to Time Series Clustering: A Simulation Study. *Journal of Classification* **27**: 333–362.
- Timmer, J., Lauk, M., Vach, W. y Lücking, C.H. (1999) A test for a difference between spectral peak frequencies. *Computational Statistics & Data Analysis*, **30**: 45–55.