

Bandas de Predicción para Datos Funcionales Basados en Medidas de Profundidad

Antonio Elías Fernández, Estudiante de Doctorado en Estadística.
Raúl Jiménez Recaredo, Departamento de Estadística.
Universidad Carlos III de Madrid.

Con el avance de las nuevas tecnologías, cada vez más y más datos son recogidos y almacenados de forma casi continua durante un intervalo de tiempo. Ejemplos donde esto sucede son las áreas de Finanzas, Medicina, Meteorología, etc. La naturaleza continua y compleja de esta información ha presentado un reto para la estadística y, por ello, la literatura ha investigado esta área denominándola Análisis de Datos Funcionales (FDA), ver Ramsay y Silverman (2005), donde se enmarca este trabajo.

Supongamos que tenemos una muestra aleatoria de funciones observadas $\{Y_1, \dots, Y_n\}$ con valores en el dominio $\mathcal{C}(I)$, $I = [a, b]$ y una función Y_{short} que es observada en un subdominio $I_{short} \subseteq I$. En este trabajo, planteamos el problema de extender la función parcialmente observada al dominio observado.

Proponemos un método no paramétrico para resolver este problema motivado por el trabajo de Sugihara y May (1990). La novedad del enfoque se centra en la selección de submuestras que hacen al dato que queremos predecir una función profunda en el rango de observación. Las regiones centrales delimitadas por las k -funciones más profundas proporcionan bandas ajustadas que envuelven el dato no sólo en la parte observada si no también conforman una banda de predicción natural en el dominio no observado.

Para resolver el problema de obtención de submuestras, presentamos dos heurísticas diseñadas para usar en conjunto con herramientas de computación y visualización de regiones centrales para datos funcionales, ver Hyndman y Shang (2010, 2015) y Sun y Genton (2012).

Finalmente, discutimos dos estudios de caso para poner en práctica nuestra metodología. Primero, predecimos la demanda eléctrica española durante los últimos tres meses de 2016 con un conjunto de funciones diarias desde 2014. Por otro lado, proponemos un ejercicio de imputación de datos perdidos con un conjunto de datos de temperaturas diarias en España, observadas en 53 estaciones ubicadas por todo el país desde 1904.

Palabras clave:

Complex Data, FDA, Depth Measures, Central Regions, Prediction.

Referencias:

- Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29{45}.
- Ramsay, J.O and Silverman, B. W. (2005). Functional Data Analysis. Springer Series in Statistics. New York: Springer.
- Shang HL, Hyndman RJ (2015). Rainbow: Rainbow Plots, Bagplots and Boxplots for Functional Data. R package version 3.4.
- Sugihara, G. and May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344(6268):734{741}.
- Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316{334}.