

Clustering de Datos Funcionales

Aplicación a Datos de Consumo Eléctrico

Andrés Castrillejo

Instituto de Estadística, FCEA. Udelar.

Fernando Massa*

Instituto de Estadística, FCEA. Udelar.

Resumen

El Análisis de Clusters es una técnica descriptiva de clasificación no supervisada que permite clasificar objetos en grupos homogéneos siguiendo un criterio de similitud. Así, los grupos obtenidos contienen objetos que son similares en su interior y disímiles entre grupos. Existe una vasta literatura, incluyendo varios algoritmos de clasificación y criterios de similitud, que tratan el caso en que los objetos son representados por un vector de medidas.

En esta aplicación se presentan los resultados de un Análisis de Cluster en el cual los objetos son representados mediante funciones. Este tipo de datos de naturaleza infinita son estudiados por la rama de la estadística conocida como Análisis de Datos Funcionales. Entre los problemas a tratar se encuentra la representación de los datos y el uso de un criterio de similitud consistente con la naturaleza de los objetos funcionales.

El trabajo está guiado por el análisis a la demanda de energía eléctrica en Uruguay. El objetivo es obtener un agrupamiento que permita describir e interpretar diferencias de comportamiento en la demanda diaria de energía eléctrica a nivel del sistema eléctrico. Los objetos con los que se trata son trayectorias de procesos estocásticos que se representan como funciones del tiempo. La no estacionariedad del proceso subyacente introduce una dificultad adicional.

Se dispone de datos de demanda de energía eléctrica muestreados cada 10 minutos sobre un período de 5 años. Los datos históricos son segmentados en días, los que conforman las unidades del análisis. La implementación propuesta implica un pre-procesamiento de los datos y su representación adecuada, para poder aproximar las funciones diarias de demanda a partir de las medidas discretas observadas, previo a la aplicación de algoritmos de clustering.

Se exploran diversas estrategias de pre-procesamiento de los datos funcionales en esta situación y se comparan diferentes algoritmos de clustering y nociones de disimilitud entre los objetos funcionales.

Palabras claves: análisis de datos funcionales; clustering ; medidas de disimilaridad

*Email: fmasa@iesta.edu.uy.