

Clustering subsecuencial de series de tiempo: Evidencia de patrones temporales en el tipo de cambio UsdMxn

Una gran cantidad de procesos naturales, financieros, tecnológicos y sociales producen grandes cantidades de datos ordenados en el tiempo, es decir, datos del tipo series de tiempo. El análisis de este tipo de datos ha captado la atención de una gran parte de la comunidad de científicos de datos y, argumentablemente, técnicas de “clustering” o “agrupamiento” son unas de las principales para el análisis de este tipo de datos.

Este trabajo es sobre Clustering SubSecuencial de Series de Tiempo, una técnica que busca agrupar subsecuencias contenidas dentro de una misma serie de tiempo, por medio del cálculo de un término de distancia euclidiana a manera de medida de similitud entre los datos. Trabajos de investigadores en la Universidad de California Riverside han mostrado la carencia de sentido en el uso de ciertas técnicas de clustering para series de tiempo, particularmente K-Means con “ventanas móviles” puesto que utilizando conjuntos de datos completamente diferentes se obtenían centroides estadísticamente similares. Los autores de UC Riverside mencionan que otros investigadores han conducido las pruebas sugeridas y han manifestado la reproducibilidad de los resultados en sus propios trabajos, los cuales han sido principalmente series de tiempo generadas por procesos biológicos y físicos. Algo a destacar es que, de acuerdo a lo escrito por los autores, estos hallazgos obtenidos invalidan los resultados de decenas de trabajos previamente publicados.

Hacemos uso del algoritmo propuesto por el grupo de investigación de UC Riverside, MASS (Mueen's Algorithm for Similarity Search), para la identificación de patrones temporales en la subsecuencia de precios intradía del tipo de cambio Dólar americano Vs Peso Mexicano (UsdMxn). La hipótesis principal es que existen condiciones bajo las cuales se puede clasificar el comunicado de un indicador macroeconómico en México o Estados Unidos como un evento disparador que genera patrones temporales en la serie de tiempo del tipo de cambio UsdMxn.

Una búsqueda quasi-exhaustiva de evidencia es conducida utilizando 10 años de información, 14.5 Millones de precios (OHLC de cada minuto), 36,000 mediciones de indicadores macroeconómicos. Debido a la cantidad de datos a procesar, el uso de un cluster de 64 núcleos ha sido planteado como parte fundamental del proyecto, tomando como ventaja que el algoritmo MASS soporta la paralelización y velocidad de convergencia de su proceso de búsqueda y ajuste.

Finalmente, con estos resultados aportamos evidencia del uso de técnicas modernas para análisis de datos del tipo series de tiempo financieras, desde un punto de estadística para el clustering subsecuencial de series de tiempo, y evidencia sobre el comportamiento de mercados financieros. Los resultados que mostramos son consistentes y documentamos las condiciones bajo las cuales, el algoritmo K-means con “ventanas móviles” produce resultados inferiores al MASS, y también, condiciones bajo las cuales no se cumple la Hipótesis del Mercado Eficiente.

Autores:

Juan Francisco Muñoz Elguezabal - franciscome@iteso.mx
Riemann Ruiz Cruz - riemannruiz@iteso.mx