

Presentación

En este siglo la revolución en la velocidad de almacenamiento y computación, de transmisión de la información y la disponibilidad de grandes bases de datos masivos, el conocido en el argot como *big data*, han abierto en la última década nuevas posibilidades en la predicción de fenómenos complejos en economía y finanzas. Desde el campo de aprendizaje automático (*Machine Learning*, *ML* por sus siglas en inglés) e inteligencia artificial, se han desarrollado nuevos métodos de predicción que, prescindiendo de hipótesis sobre el modelo generador de los datos, tratan de explotar las relaciones (posiblemente no lineales) existentes entre muchas variables en grandes bancos de datos generados automáticamente. Estos nuevos datos han impulsado avances en Estadística y Econometría en el análisis conjunto de grandes sistemas de series temporales interdependientes, con el objetivo de mejorar las predicciones. En particular, la combinación de varios métodos y modelos está proporcionando buenos resultados en gestión de carteras, análisis de coyuntura, modelos de riesgo y otras muchas áreas. Este libro tiene por objeto presentar y evaluar algunas aplicaciones en economía y empresa de los nuevos enfoques de predicción basados en datos masivos.

En sus ocho capítulos, reconocidos expertos en este campo analizarán cómo pueden mejorarse las predicciones incorporando a los enfoques tradicionales los resultados de los modelos factoriales dinámicos, árboles de decisión, bosques aleatorios (*Random Forest*) y redes neuronales, entre otros procedimientos desarrollados en los últimos años. También, se analizará como el análisis de sentimientos y de textos, así como otras nuevas herramientas no convencionales que se basan en la abundancia de datos no estructurados pero con potencial valor predictivo, pueden aportar mejoras a los métodos de predicción habitualmente utilizados. El conjunto de los capítulos publicados en este libro representa una visión amplia de aplicaciones de procedimientos de predicción ML en el contexto de variables económicas y financieras.

En el primer capítulo, **Ángela Caro** y **Daniel Peña** describen la evolución histórica de la predicción basada en series temporales y revisan algunas de las herramientas actuales más utilizadas para la predicción de datos masivos. Utilizando *redes neuronales profundas* (*Deep learning*) para la predicción de varias magnitudes macroeconómicas en 35 países de la OCDE. Estas predicciones se comparan con las obtenidas mediante modelos univariantes y modelos factoriales dinámicos con el objetivo de ilustrar las dificultades con las que se encuentra el analista en la construcción de las reglas de predicción.

En el segundo capítulo, **Christian Brownless, Jordi Llorens y Nuria Senar** justifican la necesidad de *regularización*, o imponer restricciones sobre los parámetros en la estimación, para calcular matrices de varianzas y covarianzas dinámicas en el contexto de carteras de activos financieros de alta dimensión. Se demuestra e ilustra cómo la regularización aumenta la precisión en la estimación de dichas matrices.

En el tercer capítulo, **Máximo Camacho, Salvador Ramallo y Manuel Ruiz Marín** realizan una descripción de los *árboles de decisión* en el contexto de datos independientes observados en un momento del tiempo determinado. La metodología es implementada para predecir el precio de la vivienda en Madrid. Dado el fuerte componente no lineal de la relación entre el precio y sus determinantes, la utilización de árboles de decisión es ventajosa frente a metodologías más tradicionales que pueden tener dificultades representando dichas no linealidades.

El cuarto capítulo, escrito por **Montserrat Guillen, María Láinez, Ana M. Pérez-Marín y Eduardo Sánchez**, analiza cómo los datos masivos están afectando a la medición del riesgo y la fijación de primas en seguros de automóviles. El trabajo presenta un interesante análisis sobre cómo es posible utilizar la *información de los sensores en los vehículos* para monitorizar la conducción y detectar los factores que afectan al riesgo en el contexto de los seguros de automóviles.

El capítulo quinto presenta la contribución de **Desamparados Blázquez, Josep Doménech y Ana Debón**. Estos autores utilizan un repositorio público y de libre acceso que contiene *capturas de más de 400 billones de sitios web* y analizan cómo la información sobre cambios en las páginas web de una empresa puede aportar información sobre la supervivencia empresarial.

El capítulo seis, debido a **Argimiro Arratia**, utiliza la información escrita en medios digitales que contiene información no estructurada puede combinarse con información cuantitativa tradicional para mejorar las predicciones. Se describe cómo construir un modelo de predicción con *indicadores de sentimiento* derivados de datos textuales para series temporales financieras que incluyen indicadores extraídos de noticias sobre mercados financieros.

En el capítulo siete **Corinna Guirelli, Samuel Hurtado, Javier J. Pérez y Alberto Urtasun** analizan los desafíos a los que los bancos centrales se enfrentan a la hora de utilizar datos granulares y obtenidos con mayor frecuencia de las previamente habituales, para llevar a cabo sus funciones. Los autores señalan que, en 2019, más del 60% de los bancos centrales utilizaron este tipo de datos en sus operaciones y dos tercios de ellos, los usaron como instrumento en el diseño de sus políticas. Entre los casos de éxito descritos aparecen la *medición de la incertidumbre económica basada en artículos de prensa*, el uso de informes regulares de los bancos centrales como herramienta de comunicación sobre el estado de la economía y la predicción macroeconómica.

Finalmente, **Lorenzo Pascual** y **Esther Ruiz** realizan una *descripción de la literatura sobre aplicaciones* empíricas en las que los procedimientos ML se han utilizado con éxito en la predicción económica. En comparación con procedimientos tradicionales, ML tiene mejor comportamiento predictivo en situaciones en las que las no-linealidades son importantes (como en épocas de crisis) y cuando el horizonte de predicción se incrementa.

Estos trabajos fueron presentados en una Jornada celebrada en Funcas el 8 de octubre de este año y el lector interesado puede encontrar en la web las grabaciones de las presentaciones que allí se hicieron. Los editores queremos agradecer a Funcas su apoyo en la realización de este libro que esperamos contribuya a difundir nuevas oportunidades para mejorar la predicción incorporando datos masivos.

Daniel Peña, Pilar Poncela y Esther Ruiz

Diciembre, 2020