

Prefacio

Vivimos en un mundo conectado e interdependiente. De hecho la mayoría de los elementos que nos rodean, y nosotros mismos, formamos estructuras reticulares de relaciones. Desde las redes sociales virtuales en las que participamos hasta las redes tróficas que mantienen las especies de un ecosistema, pasando por las conexiones entre diferentes entidades financieras, las redes de transporte, las estructuras moleculares,...

En estos sistemas, la suma de las partes no es igual al todo, en el sentido de que las influencias generadas por estas relaciones originan *propiedades emergentes* y comportamientos para el colectivo que estudiadas de forma individual no tendrían sentido. Por esta razón, a este tipo de redes se las denomina *sistemas complejos*. El análisis de estos sistemas permite el estudio de la propagación de las epidemias, las crisis financieras, la investigación criminológica o la comunicación y transferencia de mensajes, así como la identificación de los puntos críticos de estas estructuras y fenómenos.

La Teoría de Grafos, creada por Leonard Euler en el siglo XVIII en Königsberg (la actual Kaliningrado), ocupa un espacio relevante y sobradamente conocido en la Historia de las Matemáticas. Las aplicaciones de esta teoría en los diferentes campos científicos son difíciles de cuantificar, puesto que permite la representación de las relaciones entre objetos de todo tipo, de modo que diferentes resultados de esta teoría pueden aplicarse en cualquier sistema de relaciones.

Por otro lado, vivimos en un mundo que genera y acumula datos continuamente, y la transformación de esos datos en conocimiento útil para

la toma de decisiones en las empresas, las Administraciones Públicas, los gabinetes de ingeniería o los centros de investigación científica, se hace cada vez más necesaria; imprescindible. Sin embargo, esta cantidad creciente de datos no necesariamente se traduce en una mayor cantidad de información, ni mucho menos en un conocimiento detallado de las poblaciones que generan esos datos, ya que la estructura y la dimensión de las nuevas bases de datos dificultan su tratamiento mediante las técnicas descriptivas tradicionales.

El análisis masivo de datos, con el fin de extraer conocimiento útil de las grandes bases de datos, debe entonces plantearse desde dos vertientes entrelazadas: desde el punto de vista de las múltiples características de cada objeto bajo estudio y desde el punto de vista de las relaciones que existen entre dichos objetos. La combinación de diferentes técnicas matemáticas y computacionales nos permite extraer conocimiento tanto de las características individuales que forman el sistema como de sus propiedades colectivas, facilitando la predicción de diferentes fenómenos que afectan a diferentes grupos y permitiendo explicar la reacción de los mismos ante diferentes estímulos.

Este libro trata de exponer de forma sistemática las principales técnicas de análisis masivo de datos a estudiosos y profesionales interesados en la extracción del conocimiento de redes y grandes bases de datos, que necesitan una base teórica y práctica de Minería de Datos, Inteligencia Artificial y Teoría de Grafos para poder aplicar estas técnicas. Que por un lado necesitan comprender el funcionamiento de los algoritmos básicos para aplicarlos, adaptarlos y/o mejorarlos, y por otro lado necesitan diseñar soluciones concretas en su trabajo tomando como punto de partida aplicaciones similares implementadas en otros ámbitos. No en vano, el desarrollo sucinto de algunos casos prácticos paradigmáticos es uno de nuestros objetivos principales.

Así pues, consideramos fundamental un alto compromiso entre la teoría matemática que subyace a cada una de las técnicas y su implementación computacional en ejemplos representativos, útiles y atractivos, que pongan de manifiesto los resultados de cada técnica en la extracción del conocimiento.

El libro se compone de tres partes bien diferenciadas, pero se comienza con un capítulo de introducción dedicado a exponer, aclarar o definir algunos conceptos preliminares, a revisar el recorrido histórico de estos conceptos y a comentar algunas aplicaciones tecnológicas actuales.

En la primera parte se estudian las técnicas matemáticas y computacionales clásicas de la extracción del conocimiento en grandes bases de datos. El lector encontrará en esta primera parte un compendio detallado de los principales paradigmas de la Minería de Datos.

En la segunda parte se estudian los principales elementos de análisis de redes, como las propiedades emergentes derivadas de procesos auto-organizados, la visualización de redes, la detección de elementos críticos del sistema o la detección de comunidades y tramas más o menos aisladas.

Finalmente, aunque en cada capítulo suelen aparecer ejemplos concretos y reales de aplicación de cada técnica, en la tercera parte se presentan de forma ampliada algunas aplicaciones de las técnicas de análisis estudiadas en las dos primeras partes, tales como el diseño y desarrollo de sistemas de recomendación, modelos de análisis de riesgos en sistemas de interdependencia, técnicas de procesamiento del lenguaje natural y minería de textos, así como algunos algoritmos básicos de visión artificial y tratamiento de imágenes y sonidos.