Big data, big questions| A dozen ways to get lost in translation: Inherent challenges in large scale data sets
Big Data, Big Questions| A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets
Lawrence Busch

Abstract

As noted by the late Susan Leigh Star, technoscientific research always involves simplification and standardization. In recent years, the collection and analysis of large-scale data sets (LSDS) have become the norm. These are often convenience samples analyzed by data mining techniques. Moreover, these data are often used as the basis for public and private policy and action. At the same time, the term “large-scale” suggests completeness, while ease of collection and analysis suggest that little else need be done. Both tend to crowd out other interpretations; hence understanding their limits should be of the utmost concern. This article discusses a number of the issues of concern that arise out of the necessary but potentially problematic simplifications/standardizations found in LSDS.

Full Text:
PDF
As the data keeps getting bigger, deep learning is coming to play a key role in providing big data predictive analytics solutions. In this paper, we provide a brief overview of deep learning, and highlight current research efforts and the challenges to big data, as well as the future trends. While Big Data offers the great potential for revolutionizing all aspects of our society, harvesting of valuable knowledge from Big Data is not an ordinary task. The large and rapidly growing body of information hidden in the unprecedented volumes of non-traditional data requires both the development of advanced technologies and interdisciplinary teams working in close collaboration. To address these big data challenges, a new generation of scalable data management technologies has emerged in the last five years. Relational database management systems, which provide strong data-consistency guarantees based on vertical scaling of compute and storage hardware, are being replaced by NoSQL (variously interpreted as "No SQL", or "Not Only SQL") data stores running on horizontally-scaled commodity hardware. Designing for scale. Big data systems are inherently distributed systems. Hence, software architects must explicitly deal with issues of partial failures, unpredictable communications latencies, concurrency, consistency, and replication in the system design. Critiquing Big Data: Politics, Ethics, Epistemology | Special Section Introduction. PDF. Kate Crawford, Mary L. Gray, Kate Miltner. 10 pgs. The Big Data Divide. A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large Scale Data Sets. ABSTRACT PDF. Lawrence Busch.