Ravindra Babu / Subrahmanya / Narasimha Murty

Compression Schemes for Mining Large Datasets

A Machine Learning Perspective

This book addresses the challenges of data abstraction generation using a least number of database scans, compressing data through novel lossy and non-lossy schemes, and carrying out clustering and classification directly in the compressed domain. Schemes are presented which are shown to be efficient both in terms of space and time, while simultaneously providing the same or better classification accuracy. Features: describes a non-lossy compression scheme based on run-length encoding of patterns with binary valued features; proposes a lossy compression scheme that recognizes a pattern as a sequence of features and identifying subsequences; examines whether the identification of prototypes and features can be achieved simultaneously through lossy compression and efficient clustering; discusses ways to make use of domain knowledge in generating abstraction; reviews optimal prototype selection using genetic algorithms; suggests possible ways of dealing with big data problems using multiagent systems.

As data mining algorithms are typically applied to sizable volumes of high-dimensional data, these can result in large storage requirements and inefficient computation times. This unique text/reference addresses the challenges of data abstraction generation using a least number of database scans, compressing data through novel lossy and non-lossy schemes, and carrying out clustering and classification directly in the compressed domain. Schemes are presented which are shown to be efficient both in terms of space and time, while simultaneously providing the same or better classification accuracy, as illustrated using high-dimensional handwritten digit data and a large intrusion detection dataset. Topics and features: presents a concise introduction to data mining paradigms, data compression, and mining compressed data; describes a non-lossy compression scheme based on run-length encoding of patterns with binary valued features; proposes a lossy compression scheme that recognizes a pattern as a sequence of features and identifying subsequences; examines whether the identification of prototypes and features can be achieved simultaneously through lossy compression and efficient clustering; discusses ways to make use of domain knowledge in generating abstraction; reviews optimal prototype selection using genetic algorithms; suggests possible ways of dealing with big data problems using multiagent systems. A must-read for all researchers involved in data mining and big data, the book proposes each algorithm within a discussion of the wider context, implementation details and experimental results. These are further supported by bibliographic notes and a glossary.
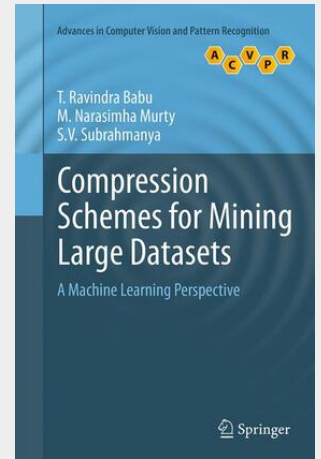
53,49 €
49,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9781447170556
Medium: Buch
ISBN: 978-1-4471-7055-6
Verlag: Springer
Erscheinungstermin: 17.09.2016
Sprache(n): Englisch
Auflage: Softcover Nachdruck of the original 1. Auflage 2013
Serie: Advances in Computer Vision and Pattern Recognition
Produktform: Kartoniert
Gewicht: 3343 g
Seiten: 197
Format (B x H): 155 x 235 mm