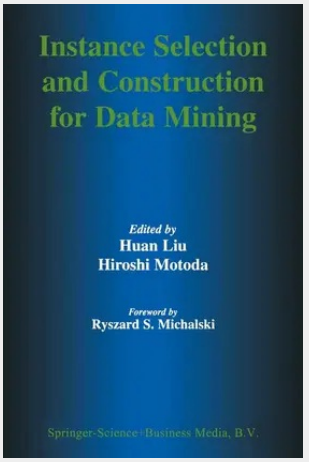
Motoda

Instance Selection and Construction for Data Mining



160,49 €
149,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

The ability to analyze and understand massive data sets lags far behind the ability to gather and store the data. To meet this challenge, knowledge discovery and data mining (KDD) is growing rapidly as an emerging field. However, no matter how powerful computers are now or will be in the future, KDD researchers and practitioners must consider how to manage ever-growing data which is, ironically, due to the extensive use of computers and ease of data collection with computers. Many different approaches have been used to address the data explosion issue, such as algorithm scale-up and data reduction. Instance, example, or tuple selection pertains to methods or algorithms that select or search for a representative portion of data that can fulfill a KDD task as if the whole data is used. Instance selection is directly related to data reduction and becomes increasingly important in many KDD applications due to the need for processing efficiency and/or storage efficiency.
One of the major means of instance selection is sampling whereby a sample is selected for testing and analysis, and randomness is a key element in the process. Instance selection also covers methods that require search. Examples can be found in density estimation (finding the representative instances - data points - for a cluster); boundary hunting (finding the critical instances to form boundaries to differentiate data points of different classes); and data squashing (producing weighted new data with equivalent sufficient statistics). Other important issues related to instance selection extend to unwanted precision, focusing, concept drifts, noise/outlier removal, data smoothing, etc. brings researchers and practitioners together to report new developments and applications, to share hard-learned experiences in order to avoid similar pitfalls, and to shed light on the future development of instance selection. This volume serves as a comprehensive reference for graduate students, practitioners and researchers in KDD.

The ability to analyze and understand massive data sets lags far behind the ability to gather and store the data. To meet this challenge, knowledge discovery and data mining (KDD) is growing rapidly as an emerging field. However, no matter how powerful computers are now or will be in the future, KDD researchers and practitioners must consider how to manage ever-growing data which is, ironically, due to the extensive use of computers and ease of data collection with computers. Many different approaches have been used to address the data explosion issue, such as algorithm scale-up and data reduction. Instance, example, or tuple selection pertains to methods or algorithms that select or search for a representative portion of data that can fulfill a KDD task as if the whole data is used. Instance selection is directly related to data reduction and becomes increasingly important in many KDD applications due to the need for processing efficiency and/or storage efficiency. One of the major means of instance selection is sampling whereby a sample is selected for testing and analysis, and randomness is a key element in the process. Instance selection also covers methods that require search. Examples can be found in density estimation (finding the representative instances - data points - for a cluster); boundary hunting (finding the critical instances to form boundaries to differentiate data points of different classes); and data squashing (producing weighted new data with equivalent sufficient statistics). Other important issues related to instance selection extend to unwanted precision, focusing, concept drifts, noise/outlier removal, data smoothing, etc. Instance Selection and Construction for Data Mining brings researchers and practitioners together to report new developments and applications, to share hard-learned experiences in order to avoid similar pitfalls, and to shed light on the future development of instance selection. This volume serves as a comprehensive reference for graduate students, practitioners and researchers in KDD.

Artikelnummer: 9780792372097
Medium: Buch
ISBN: 978-0-7923-7209-7
Verlag: Springer
Erscheinungstermin: 28.02.2001
Sprache(n): Englisch
Auflage: Erscheinungsjahr 2001
Serie: The Springer International Series in Engineering and Computer Science
Produktform: Gebunden
Gewicht: 1780 g
Seiten: 416
Format (B x H): 160 x 241 mm

Kundenservice Verlag Dr. Otto Schmidt KG
Neumannstraße 10, 40235 Düsseldorf | kundenservice@otto-schmidt.de | 0221-93738-997 (Inland)