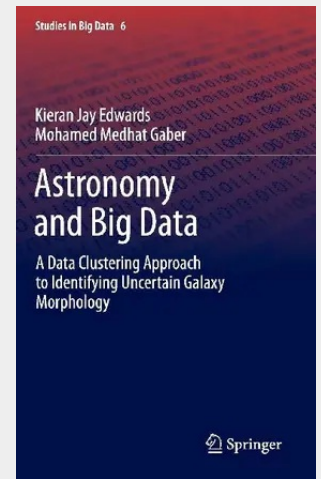


Astronomy and Big Data

With the onset of massive cosmological data collection through media such as the Sloan Digital Sky Survey (SDSS), galaxy classification has been accomplished for the most part with the help of citizen science communities like Galaxy Zoo. Seeking the wisdom of the crowd for such Big Data processing has proved extremely beneficial. However, an analysis of one of the Galaxy Zoo morphological classification data sets has shown that a significant majority of all classified galaxies are labelled as "Uncertain".

This book reports on how to use data mining, more specifically clustering, to identify galaxies that the public has shown some degree of uncertainty for as to whether they belong to one morphology type or another. The book shows the importance of transitions between different data mining techniques in an insightful workflow. It demonstrates that Clustering enables to identify discriminating features in the analysed data sets, adopting a novel feature selection algorithms called Incremental Feature Selection (IFS). The book shows the use of state-of-the-art classification techniques, Random Forests and Support Vector Machines to validate the acquired results. It is concluded that a vast majority of these galaxies are, in fact, of spiral morphology with a small subset potentially consisting of stars, elliptical galaxies or galaxies of other morphological variants.

With the onset of massive cosmological data collection through media such as the Sloan Digital Sky Survey (SDSS), galaxy classification has been accomplished for the most part with the help of citizen science communities like Galaxy Zoo. Seeking the wisdom of the crowd for such Big Data processing has proved extremely beneficial. However, an analysis of one of the Galaxy Zoo morphological classification data sets has shown that a significant majority of all classified galaxies are labelled as „Uncertain“. This book reports on how to use data mining, more specifically clustering, to identify galaxies that the public has shown some degree of uncertainty for as to whether they belong to one morphology type or another. The book shows the importance of transitions between different data mining techniques in an insightful workflow. It demonstrates that Clustering enables to identify discriminating features in the analysed data sets, adopting a novel feature selection algorithms called Incremental Feature Selection (IFS). The book shows the use of state-of-the-art classification techniques, Random Forests and Support Vector Machines to validate the acquired results. It is concluded that a vast majority of these galaxies are, in fact, of spiral morphology with a small subset potentially consisting of stars, elliptical galaxies or galaxies of other morphological variants.



106,99 €

99,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9783319065984

Medium: Buch

ISBN: 978-3-319-06598-4

Verlag: Springer International Publishing

Erscheinungstermin: 29.04.2014

Sprache(n): Englisch

Auflage: 2014

Serie: Studies in Big Data

Produktform: Gebunden

Gewicht: 354 g

Seiten: 105

Format (B x H): 160 x 241 mm

