1st ed. 2019, XXVII, 165 p. 23 illus., 21 illus. in color.

Laith Mohammad Qasim Abualigah

Feature Selection and Enhanced Krill Herd Algorithm for Text Document Clustering

Reihe: Studies in Computational Intelligence

- **Presents a new method for solving the text document clustering problem and demonstrates that it can outperform other comparable methods**
- **Covers the main text clustering preprocessing steps and the metaheuristics needed in order to deal with the text document clustering problems**
- **Proposes methods that can be applied to a broad range of text documents (e. g. newsgroup documents appearing on newswires, Internet web pages, and hospital information), modern applications (technical reports and university data), and the biomedical sciences (large biomedical datasets)**

This book puts forward a new method for solving the text document (TD) clustering problem, which is established in two main stages: (i) A new feature selection method based on a particle swarm optimization algorithm with a novel weighting scheme is proposed, as well as a detailed dimension reduction technique, in order to obtain a new subset of more informative features with low-dimensional space. This new subset is subsequently used to improve the performance of the text clustering (TC) algorithm and reduce its computation time. The k-mean clustering algorithm is used to evaluate the effectiveness of the obtained subsets. (ii) Four krill herd algorithms (KHAs), namely, the (a) basic KHA, (b) modified KHA, (c) hybrid KHA, and (d) multi-objective hybrid KHA, are proposed to solve the TC problem; each algorithm represents an incremental improvement on its predecessor. For the evaluation process, seven benchmark text datasets are used with different characterizations and complexities.