

Data Mining with Decision Trees Theory and Applications
2nd Edition (Series in Machine Perception and Artificial Intelligence, volume 81)
Authored by Lior Rokach & Oded Maimon
World Scientific Pub Co
New Jersey, London, Singapore, Beijing, Shanghai, Hong Kong, Taipei, Chennai
2015
305 pp.
100 USD hard cover
ISBN 978-9-814-59007-X

With the importance of exploring large and complex data sets in knowledge discovery and data mining, the application of decision trees have become a powerful and popular approach. Whilst the first edition focused on using trees for classification tasks, this second edition describes how decision trees can be used for regression, clustering and survival analysis – very important topics for the discovery of useful patterns in complex data sets. The first edition by Rokach & Maimon was already a classic on the desks of scientists, and in this new edition all chapters have been revised and new topics include: cost-sensitive active learning, learning with uncertain and imbalanced data, privacy preserving decision tree learning, lessons learned from comparative studies, and learning decision trees for Big Data and a entire chapter on recommender systems. Very important to note is the practical walk-through guide to existing open-source data mining software – which alone constitutes a huge additional benefit.

The book starts with a easy to read introduction to decision trees, and quickly continues to answer the question on how to train decision trees; chapter 3 introduces a generic algorithm for top-down induction of decision trees and chapter 4 contains evaluation methods. Splitting criteria and pruning trees are discussed in chapter 5 and 6, and continued by popular decision trees induction algorithms (ID3, C4.5, CART, CHAID, QUEST, etc.). Chapter 8 is dealing beyond classification tasks, i.e. regression trees, survival trees, clustering trees and hidden Markov model trees. Chapter 9 deals extensively with decision forests, including e.g. Naïve Bayes, entropy weighting and random forests. Chapter 10 is on Weka and R, whilst chapter 11 is on advanced decision trees including oblivious decision trees, lazy trees, option trees, etc. Chapter 12 deals with cost-sensitive active and proactive learning and Chapter 13 concentrates on feature selection. In chapter 14 the authors describe fuzzy decision trees and in Chapter 15 on hybridization of decision trees with other techniques, e.g. CPOM, and evolutionary algorithms. Finally, Chapter 16 deals with the use of decision trees for recommending items and preferences elicitation.

Overall, the book by Rokach & Maimon is a must read for anybody who is working in the area of knowledge discovery/data mining and has the potential to become a standard on our desks.

Andreas Holzinger
Graz