

Advances in Feature Selection for Data and Pattern Recognition

Konstantinos Konstantinidis AEM:2546

March 2021

Summary

This book addresses the uses in the field of feature selection for data and pattern recognition. Feature selection provides methods and algorithms, in numerous applications that the book indicates. The book consists of four parts:

Nature and Representation of Data

Chapter 2 presents how to take a large quantity of numerical data and reducing it during discretisation. The problem is how such reduction of data sets changes the error rate measured by the C4.5 decision tree generation system with cross-validation. The experiments show that for a Dominant Attribute discretisation method the error has much worse results but simpler decision trees.

Chapter 3 introduces the weka workbench, a graphical interface that is used for data analysis without the use of coding tools (ex. R, python). So it is easier for the data analyst to examine them in the first place.

Chapter 4 presents the motivations and main technicalities involved in the process of constructing AbDGs (Attribute-Based Decision Graphs), also stresses showing some of the pros of this graph-based structure (ex. robustness and low computational costs) associated with both, training and memory use.

Chapter 5 is focused on the changes of an extension of dynamic programming approach for optimization of decision rules relative to length. The chapter also provides results of experiments from UCI Machine Learning Repository.

Ranking and exploration of features

Chapter 6 focuses on generational Feature Elimination and Ranking Feature Selection Methods. It shows an overview of reasons for using ranking feature selection methods and the main general classes of this kind of algorithms are described. Also, experiments are performed on real-life data sets.

Chapter 7 explains the Ranking-Based Rule Classifier Optimisation. Ranking is a strategy widely used for estimating relevance or importance of available characteristic features. The chapter describes research on the latter approach,

involving filtering inferred decision rules while exploiting ranking positions and scores of features.

Chapter 8 addresses the method for attribute selection in a dispersed decision-making system. Dispersed knowledge is understood to be the knowledge that is stored in the form of several decision tables. Different methods for solving the problem of classification based on dispersed knowledge are considered. The results obtained for five data sets from the UCI Repository are compared.

Chapter 9 contains the study of knowledge representation in rule-based knowledge bases. Feature selection is discussed as a part of mining knowledge bases from a knowledge engineer's and from a domain expert's perspective. The objective of variable selection has many provides numerous benefits such as improving prediction performance, faster and more cost-effective predictors.

Image, shape, motion, and audio detection and recognition

Chapter 10 explores recent advances in brain imaging technology, in addition with large-scale brain research projects. It allows to capture brain activity in unprecedented detail. In principle, the observed data is expected to substantially shape the knowledge about brain activity, which includes the development of new biomarkers of brain disorders.

Chapter 11 introduces the notion of classes of shapes that have descriptive proximity to each other in planar digital 2D image object shape detection. The research is focused on the triangulation of image object shapes, resulting in maximal nerve complexes from which shape contours and shape interiors can be detected and described.

Chapter 12 shows an experimental study of several methods for real motion intent classification. Methods comparison and obtained results are given, and a study of features feeding the classifiers is provided. Differences among participating subjects and accuracies for real and imaginary motion are discussed.

Chapter 13 is an extension of the work presented where the problem of classifying audio signals using a supervised tolerance class learning algorithm (TCL) based on tolerance near sets was first proposed. Extensive experimentation with different audio-video data sets was performed on TCL algorithm.

Recognition and decision support systems

Chapter 14 overviews an application area of recommendations for customer loyalty improvement, which has become a very popular and important topic area in today's business decision problems. Major machine learning techniques used to develop knowledge-based recommender system, such as decision reducts, classification, clustering, action rules, are described.

Chapter 15 presents a discussion on an alternative attempt to manage the grids that are in intelligent buildings such as central heating, heat recovery ventilation or air conditioning for energy cost minimization. A suggested matrix-like grid that includes methods for achieving the expected minimization goals is also presented.