

Studies in Fuzziness and Soft Computing

Volume 416

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

The series “Studies in Fuzziness and Soft Computing” contains publications on various topics in the area of soft computing, which include fuzzy sets, rough sets, neural networks, evolutionary computation, probabilistic and evidential reasoning, multi-valued logic, and related fields. The publications within “Studies in Fuzziness and Soft Computing” are primarily monographs and edited volumes. They cover significant recent developments in the field, both of a foundational and applicable character. An important feature of the series is its short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by SCOPUS, DBLP, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at <https://link.springer.com/bookseries/2941>

Mahdi Eftekhari · Adel Mehrpooya ·
Farid Saberi-Movahed · Vicenç Torra

How Fuzzy Concepts Contribute to Machine Learning

Mahdi Eftekhari
Department of Computer Engineering
Shahid Bahonar University of Kerman
Kerman, Iran

Adel Mehrpooya
Department of Computer Engineering
Shahid Bahonar University of Kerman
Kerman, Iran

Farid Saberi-Movahed
Department of Applied Mathematics
Graduate University of Advanced
Technology
Kerman, Iran

Vicenç Torra
Department of Computing Sciences
Umeå University
Umeå, Sweden

ISSN 1434-9922

ISSN 1860-0808 (electronic)

Studies in Fuzziness and Soft Computing

ISBN 978-3-030-94065-2

ISBN 978-3-030-94066-9 (eBook)

<https://doi.org/10.1007/978-3-030-94066-9>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To the soul of my father who taught like a loving teacher until the last moments of his life and dedicated to my mother a living statue of love.

Mahdi Eftekhari

To my parents with love, who gave me “95% help and support” and “95% mental health issues” to enable me to handle this formidable task. No, it did not make up 190%, they multitasked.

Adel Mehrpooya

To Parvin Rezaei who opened up new horizons in my life.

Farid Saberi-Movahed

To the researchers that have been studying HFS.

Vicenç Torra

Preface

The aim of this book is to introduce some new trends of research regarding the use of traditional fuzzy sets, particularly hesitant fuzzy sets, in machine learning. In data-driven systems, we often need to deal with uncertainty. This uncertainty is caused by the available data and also in stochastic elements inherent to the problems being modeled. Machine learning algorithms face these problems when building the models. Fuzzy set theory provides some tools for this purpose. In this book, we review some of the tools based on hesitant fuzzy sets that can be used in machine learning problems. This book is addressed to audiences who are interested in both machine learning and fuzzy set extensions.

The contents are organized into three parts including ten chapters in addition to a chapter on preliminaries which is an introduction to various concepts used in the next three parts. That is, Chap. 1 presents the preliminaries.

Part I focuses on the application of both fuzzy set and hesitant fuzzy set concepts in clustering algorithms and on other unsupervised learning approaches. This part begins with Chap. 2 that describes the application of hesitant fuzzy concepts to fuse the results of different clustering algorithms. This chapter also presents the concept of hesitant fuzzy partitions. Then, Chap. 3 introduces an unsupervised feature selection method based on the concepts of sensitivity and correlation. The definition of sensitivity given in this chapter is based on the gradient of density function in subtractive clustering with respect to a given feature.

Part II discusses supervised learning problems and explains cases for which fuzzy and hesitant fuzzy concepts can be used to boost the performance of supervised tasks. It is composed of Chaps. 4–7 whose description is briefly provided here. Chapters 4 and 5 study two extensions of fuzzy decision trees developed in recent years. Chapter 6 focuses on using hesitant fuzzy sets in decision trees when the data to learn the trees are imbalanced. That is, the use of fuzzy sets for imbalanced classification problems, a particular type of supervised learning tasks. More precisely, the chapter describes the use of various information gain measures to combine information, as well as the use of concepts related to hesitant fuzzy sets to combine the results of fuzzy decision trees. Then, Chap. 7 discusses the application of hesitant fuzzy sets to ensemble learning algorithms. In ensemble learning, and more particularly in dynamic ensemble selection problems, situations arise in which multiple-criteria

decision-making notions can be used. Chapter 7 concludes this part. It considers the problem of considering different machine learning algorithms as a set of experts and how to process the associated information using hesitant fuzzy elements.

Part III provides a brief survey of recent uses of hesitant fuzzy set and rough set concepts in supervised dimension reduction problems. In Chap. 8, some similarity measures as well as feature evaluation metrics are defined in form of hesitant fuzzy sets. Then these sets are applied to combine different measures and criteria of feature selection. The main idea of this chapter is to introduce some methods for combining different feature ranking algorithms and metrics via hesitant fuzzy sets. A distributed version of hesitant fuzzy-based algorithms is then provided in Chap. 9. This distributed version is appropriate for handling big data problems. Then, Chap. 10 reviews an approach to combine different rough set-based feature selection metrics through hesitant fuzzy sets. This part is concluded with Chap. 11. In this last chapter, we explain how hesitant fuzzy correlation can be used to tackle supervised feature selection issues.

All the approaches presented in this book have a common motif. They use hesitant fuzzy set concepts in machine learning problems. They consider these problems from a multi-criteria decision-making perspective. More particularly, these problems are framed considering different machine learning algorithms as experts, and then taking the results of these algorithms as expert's opinions. Then, hesitant fuzzy sets are used to combine these opinions. The algorithms presented in this book have been described and tested in a set of papers. These papers report the good performance of the approaches.

In data-driven machine learning, model building and selection is based on a set of criteria. Naturally, efficiency and accuracy is one of them. This usually implies that more complex models with more parameters perform better. This contrasts with model simplicity (and Occam's razor, as we mention in Chap. 5.1). Explainable AI, the need to build transparent models and make transparent decisions add additional constraints to how machine learning models are identified and how these models are used. Privacy regulations also add additional constraints to model building and selection. The need to consider all these aspects into account make machine learning problems challenging. In this book, we have described some of the first contributions of hesitant fuzzy sets for this complex problem. Further solutions can be developed to tackle with these other competing requirements.

The authors appreciate and value the work of students who have contributed to the papers used as the major references of this book over the course of 6 years since 2014. The authors gratefully acknowledge Ms. S. Barchinejad, Ms. L. Aliahmadipour, Ms. M. Mokhtia, Ms. S. Sardari, Mr. M. Zeinalkhani, Mr. M. Mohtashami, Mr. M. K. Ebrahimpour, and Mr. J. Elmi.

Kerman, Iran
Kerman, Iran
Kerman, Iran
Umeå, Sweden
December 2021

Mahdi Eftekhari
Adel Mehrpooya
Farid Saberi-Movahed
Vicenç Torra

Contents

1	Preliminaries	1
1.1	Fuzzy Notions	1
1.1.1	Fuzzy Sets	1
1.1.2	Membership Functions	2
1.1.3	Hesitant Fuzzy Sets	5
1.2	Decision Tree	12
1.2.1	Fuzzy Decision Tree	13
1.3	Dimensionality Reduction and Related Subjects	20
1.3.1	Pearson's Correlation Coefficient Measure	21
1.3.2	Correlation Coefficient of Hesitant Fuzzy Sets	22
1.3.3	Correlation-Based Merit	23
1.3.4	Similarity Measures	24
1.3.5	Rough Set and Fuzzy-Rough Set Basic Concepts	25
1.3.6	Weighted Rough Set Basic Concepts	27
1.3.7	Fuzzy Rough Set Basic Concepts	27
1.4	Fuzzy Clustering	29
1.4.1	Fuzzy c -Means Clustering	29
1.4.2	Subtractive Clustering	30
1.4.3	Fuzzy Partitions	31
1.4.4	I-Fuzzy Partitions	32
1.5	Linear Regression	33
1.5.1	Ridge Regression	34
1.5.2	LASSO Regression	34
1.5.3	Elastic Net Regression	34
	References	35
 Part I Unsupervised Learning		
2	A Definition for Hesitant Fuzzy Partitions	41
2.1	Introduction	41
2.2	H-Fuzzy Partition	42

2.2.1	Construction of H-Fuzzy Partitions	43
2.3	Discussion	49
2.4	Computer Programming Exercises for Future Works	49
	References	49
3	Unsupervised Feature Selection Method Based on Sensitivity and Correlation Concepts for Multiclass Problems	51
3.1	Introduction	51
3.1.1	GA for Feature Selection	52
3.2	Proposed Unsupervised Feature Selection Method	53
3.2.1	Feature Relevance Evaluation via Sensitivity Analysis Based on Subtractive Clustering	53
3.2.2	A General Scheme for Sensitivity and Correlation Based Feature Selection (SCFS)	54
3.3	Discussion	55
3.4	Computer Programming Exercise for Future Works	57
	References	57
 Part II Supervised Learning Classification and Regression		
4	Fuzzy Partitioning of Continuous Attributes Through Crisp Discretization	61
4.1	Introduction	61
4.2	Discretization Methods	62
4.3	Employing Discretization Methods for Fuzzy Partitioning	64
4.3.1	Defining MFs Over Crisp Partitions	66
4.3.2	Fuzzy Entropy Based Fuzzy Partitioning	75
4.4	Discussion	77
4.5	Computer Programming Exercises for Future Works	78
	References	79
5	Comparing Different Stopping Criteria for Fuzzy Decision Tree Induction Through IDFID3	81
5.1	Introduction	81
5.2	Stopping Criteria	82
5.3	Iterative Deepening FID3	84
5.3.1	The Stopping Criterion of IDFID3	87
5.3.2	Comparison Method for Various Stopping Criteria	89
5.4	Discussion	91
5.5	Computer Programming Exercise for Future Works	92
	References	92
6	Hesitant Fuzzy Decision Tree Approach for Highly Imbalanced Data Classification	93
6.1	Introduction	93
6.2	Hesitant Fuzzy Decision Tree Approach	96
6.2.1	Data Balancing	96

6.2.2	Generating the Membership Functions	97
6.2.3	Construction of Fuzzy Decision Trees	97
6.2.4	The Aggregation of FDTs	98
6.2.5	Notations for Different FDT Classifiers	98
6.3	Discussion	101
6.4	Computer Programming Exercise for Future Works	104
	References	105
7	Dynamic Ensemble Selection Based on Hesitant Fuzzy	
	Multiple Criteria Decision-Making	107
7.1	Introduction	107
7.2	Multiple Classifier Systems and Dynamic Selection	
	Techniques	108
7.2.1	Neighborhood Selection Techniques	110
7.2.2	Competence Level Calculation	110
7.2.3	Classifier Selection	113
7.3	Dynamic Ensemble Selection Based on Hesitant Fuzzy	
	Concepts	114
7.4	Discussion	115
7.5	Computer Programming Exercises for Future Works	116
	References	116
	Part III Dimension Reduction	
8	Ensemble of Feature Selection Methods: A Hesitant Fuzzy Set	
	Based Approach	121
8.1	Introduction	121
8.2	The MRMR-HFS Method	123
8.3	Discussion	130
8.4	Computer Programming Exercises for Future Works	131
	References	131
9	Distributed Feature Selection: A Hesitant Fuzzy Correlation	
	Concept for High-Dimensional Microarray Datasets	133
9.1	Introduction	133
9.2	Distributed HCPF Feature Selection Algorithm	136
9.2.1	The Partitioning Process in the Distributed	
	Version of HCPF	138
9.2.2	HCPF: A Feature Selection Algorithm Based	
	on HFS	138
9.2.3	The Merging Process in the Distributed Version	
	of HCPF	144
9.3	Discussion	144
9.4	Computer Programming Exercises for Future Works	146
	References	146

10	A Hybrid Filter-Based Feature Selection Method via Hesitant Fuzzy and Rough Sets Concepts	147
10.1	Introduction	147
10.2	Proposed Hybrid Filter-Based Feature Selection Method	149
10.2.1	Sample Weighting Approach	150
10.2.2	Primary Feature Subset Selection	150
10.2.3	Fuzzy Rough-Based Elimination	151
10.3	Discussion	155
10.4	Computer Programming Exercises for Future Works	155
	References	156
11	Feature Selection Based on Regularization of Sparsity Based Regression Models by Hesitant Fuzzy Correlation	157
11.1	Introduction	157
11.2	Feature Selection Method Based on Regularization	159
11.2.1	Regularized Ridge Regression	160
11.2.2	Regularized LASSO and Elastic Net Regression	161
11.3	Alternative Sparsity Based Regression Methods	163
11.4	Discussions	164
11.5	Computer Programming Exercises for Future Works	166
	References	166