# Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications

Muhammad Summair Raza ·
Usman Qamar

# Understanding and Using Rough Set Based Feature Selection: Concepts, Techniques and Applications

Second Edition

Springer

Muhammad Summair Raza
Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

Usman Qamar
Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering
National University of Sciences and Technology (NUST)
Islamabad, Pakistan

ISBN 978-981-32-9165-2 ISBN 978-981-32-9166-9 (eBook)
https://doi.org/10.1007/978-981-32-9166-9



This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*This book is dedicated to the memory of Prof. Zdzisław Pawlak (1926–2006), the father of Rough Set Theory. He was the founder of the Polish school of Artificial Intelligence and one of the pioneers in Computer Engineering and Computer Science.*

# Preface

Rough set theory, proposed in 1982 by Zdzislaw Pawlak, is in constant development. It is concerned with the classification and analysis of imprecise or uncertain information and knowledge. It has become a prominent tool for data analysis. This book provides a comprehensive introduction to rough set-based feature selection. It enables the reader to systematically study all topics of rough set theory (RST) including preliminaries, advance concepts, and feature selection using RST. This book is supplemented with RST-based API library that can be used to implement several RST concepts and RST-based feature selection algorithms.

The book is intended to provide an important reference material for students, researchers and developers working in the areas of feature selection, knowledge discovery and reasoning with uncertainty, especially for those who are working in RST and granular computing. The primary audience of this book is the research community using rough set theory (RST) to perform feature selection (FS) on large-scale datasets in various domains. However, any community interested in feature selection such as medical, banking, finance can also benefit from the book.

The second edition of the book now also covers dominance-based rough set approach and fuzzy rough sets. Dominance-based rough set approach (DRSA) is an extension to the conventional rough set approach which supports the preference order using dominance principle. Fuzzy rough sets are fuzzy generalization of rough sets. API library of dominance-based rough set approach is also provided with the second edition of the book.

Islamabad, Pakistan

Muhammad Summair Raza
Usman Qamar

# Contents

# About the Authors

**Muhammad Summair Raza** has Ph.D. specialization in software engineering from the National University of Sciences and Technology (NUST), Pakistan. He completed his MS from International Islamic University, Pakistan, in 2009. He is also associated with the Virtual University of Pakistan as assistant professor. He has published various papers in international-level journals and conferences with a focus on rough set theory. His research interests include feature selection, rough set theory, trend analysis, software architecture, software design and non-functional requirements.

**Dr. Usman Qamar** is currently a tenured associate professor at National University of Sciences and Technology (NUST) having over 15 years of experience in data engineering and decision sciences both in academia and industry having spent nearly 10 years in the UK. He has a Masters in Computer Systems Design from University of Manchester Institute of Science and Technology (UMIST), UK. His M.Phil. in Computer Systems was a joint degree between UMIST and University of Manchester which focused on feature selection in big data. In 2008/09, he was awarded Ph.D. from University of Manchester, UK. His Ph.D. specialization is in Data Engineering, Knowledge Discovery and Decision Science. His Post Ph.D. work at University of Manchester, involved various research projects including hybrid mechanisms for statistical disclosure (feature selection merged with outlier analysis) for Office of National Statistics (ONS), London, UK, churn prediction for Vodafone UK and customer profile analysis for shopping with the University of Ghent, Belgium. He has also done a post graduation in Medical and Health Research, from University of Oxford, UK, where he worked on evidence-based health care, thematic qualitative data analysis and healthcare innovation and technology. He is director of Knowledge and Data Science Research Centre, a Centre of Excellence at NUST, Pakistan and principal investigator of Digital Pakistan Lab, which is part of National Centre for Big Data and Cloud Computing. He has authored over 150 peer reviewed publications which includes 2 books published by Springer & Co. He has successfully supervised 5 Ph.D. students and over 70 master students. Dr. Usman has been able to acquire nearly PKR 100 million in research grants. He has received multiple research awards, including Best Researcher of

Pakistan 2015/16 by Higher Education Commission (HEC), Pakistan as well as gold in Research and Development category by Pakistan Software Houses Association (P@SHA) ICT Awards 2013 and 2017 and Silver award in APICTA (Asia Pacific ICT Alliance Awards) 2013 in category of R&D hosted by Hong Kong. He is also recipient of the prestigious Charles Wallace Fellowship 2016/17 as well as British Council Fellowship 2018, visiting research fellow at Centre of Decision Research, University of Leeds, UK and visiting senior lecturer at Manchester Metropolitan University, UK. Finally, he has the honour of being the finalist of the British Council's Professional Achievement Award 2016/17.