Advances in Intelligent Systems and Computing 1409

Thomas Jansen
Richard Jensen
Neil Mac Parthaláin
Chih-Min Lin *Editors*

Advances in Computational Intelligence Systems

Contributions Presented at the 20th UK Workshop on Computational Intelligence, September 8-10, 2021, Aberystwyth, Wales, UK

Springer

Advances in Intelligent Systems and Computing

Volume 1409

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by DBLP, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST).

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

More information about this series at http://www.springer.com/series/11156

Thomas Jansen · Richard Jensen ·
Neil Mac Parthaláin · Chih-Min Lin
Editors

Advances in Computational Intelligence Systems

Contributions Presented at the 20th UK Workshop on Computational Intelligence, September 8–10, 2021, Aberystwyth, Wales, UK

Springer

Editors
Thomas Jansen
Department of Computer Science
Aberystwyth University
Aberystwyth, UK

Richard Jensen
Department of Computer Science
Aberystwyth University
Aberystwyth, UK

Neil Mac Parthaláin
Department of Computer Science
Aberystwyth University
Aberystwyth, UK

Chih-Min Lin
Department of Electrical Engineering
Yuan Ze University
Taoyuan, Taiwan

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-030-87093-5 ISBN 978-3-030-87094-2 (eBook)
https://doi.org/10.1007/978-3-030-87094-2

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

Preface

This volume contains the papers presented at the 20th UK Workshop on Computational Intelligence (UKCI 2021), held virtually by Aberystwyth University, September 8–10, 2021. This marks the 20th anniversary of UKCI, a testament to the increasing role and importance of computational intelligence (CI) and the continuing interest in its development. UKCI provides a forum for the academic community and industry to share ideas and experience in this field. UKCI 2021 is the only event in the workshop series to be held online since the inaugural event in 2001, due to the COVID-19 pandemic. EDMA 2021, the 4th International Engineering Data- and Model-Driven Applications Workshop, is also incorporated and held in conjunction with UKCI 2021.

Paper submissions were invited in the areas of fuzzy systems, neural networks, evolutionary computation, machine learning, data mining, cognitive computing, intelligent robotics, hybrid methods, deep learning, and applications of CI. In addition to UK institutions, UKCI 2021 attracted submissions from Canada, China, India, Italy, Japan, Thailand, and the USA, reflecting the global appeal of research in this area.

Fifty papers were accepted at the workshop, with each paper having been reviewed by at least two members of the program committee. Of these, 42 were accepted for presentation at UKCI 2021 (34 as regular papers and eight as short papers). Eight papers were accepted for presentation at EDMA. The papers in this volume have been divided into eight sections: fuzzy systems, evolutionary algorithms, neural networks/deep learning, intelligent robotics, data mining/machine learning, image analysis, health informatics, and engineering data- and model-driven applications.

There were three keynote talks given by established researchers in the field: Prof. Bernadette Bouchon-Meunier (CNRS National Center for Scientific Research), Prof. Emma Hart (Edinburgh Napier University), and Dr. Una-May O'Reilly (MIT Computer Science and Artificial Intelligence Lab). We would like to thank everyone who contributed to the success of UKCI 2021. We particularly thank the members of the program committee for their reviews and recommendations, the keynote speakers, the authors and presenters, and the organizing committee. We are also grateful for the support provided by Aberystwyth University and Springer.

Neil Mac Parthaláin
Richard Jensen
Thomas Jansen
Qiang Shen

UKCI 2021
20TH UK WORKSHOP ON COMPUTATIONAL INTELLIGENCE

Contents

Fuzzy Systems

An Evolving Feature Weighting Framework for Granular Fuzzy Logic Models 3
Muhammad Zaiyad Muda and George Panoutsos

Fuzzy Multi-Criteria Decision-Making: Example of an Explainable Classification Framework 15
Hesham Yusuf, Kai Yang, and George Panoutsos

Rough-Fuzzy Rule Interpolation for Data-Driven Decision Making 27
Chengyuan Chen and Qiang Shen

Fuzzy Intelligent System for Improving High Availability of a Docker-Based Hybrid Cloud Infrastructure 38
Nitin Naik

A Study of Neuro-Weighted Nearest-Neighbour Classification 51
Qianyi Zhang, Guanli Yue, Yanpeng Qu, and Ansheng Deng

Q-Routing with Multiple Soft Requirements 61
Trevor Martin, Douglas Harewood-Gill, and Reza Nejabati

Evolutionary Algorithms

Avoiding Excess Computation in Asynchronous Evolutionary Algorithms 71
Eric O. Scott, Mark Coletti, Catherine D. Schuman, Bill Kay, Shruti R. Kulkarni, Maryam Parsa, and Kenneth A. De Jong

Stochasticity Improves Evolvability in Artificial Gene Regulatory Networks 83
Alexander Turner and Katharina C. Wollenberg Valero

ACO Inspired GA Mutation Applied to the TSP 95
Darren M. Chitty

Centralised and Decentralised Control of Video Game Agents 108
Sam G. Robinson, Chloe M. Barnes, and Peter R. Lewis

An Enhancement of the NSGA-II Reliability Optimization Using Extended Kalman Filter Based Initialization 121
Savas Yuce and Ke Li

Neural Networks/Deep Learning

Discrete-Time Recurrent Neural Network for Solving Multi-linear $\mathcal{M}$-tensor Equation 131
Huanmei Wu, Shuqiao Wang, Xiujuan Du, and Mei Liu

An Approach to Implementing Convolutional Neural Network Based on Low Density FPGA 144
Hanzhong Zhang, Qihuan Li, Xin Bai, Zhengxia Wang, and Lu Lou

Towards a Framework for Interpretation of CNN Results with ANFIS 153
Muhammad Ismail, Changjing Shang, and Qiang Shen

Syntactic Dependency Constraint Based Graph Convolutional Network for Aspect Level Sentiment Classification 167
Jinpeng You, Dazhen Lin, Donglin Cao, and Fei Chao

Faster RCNN Hyperparameter Selection for Breast Lesion Detection in 2D Ultrasound Images 179
Anu Bose, Tuan Nguyen, Hongbo Du, and Alaa AlZoubi

Asymmetric Convolution View Adaptation Networks for Skeleton-Based Human Action Recognition 191
Tianyu Ma, Jiahui Yu, Hongwei Gao, and Zhaojie Ju

Intelligent Robotics

Ankle Variable Impedance Control for Humanoid Robot Upright Balance Control 203
Kaiyang Yin, Yaxu Xue, Yifei Wang, and Longzhi Yang

Robust Control for Autonomous Surveillance with Unmanned Aerial Manipulator 215
Weiwei Zhan, Yanjie Chen, Yaonan Wang, Fei Chao, and Qiang Shen

Design and Implementation of Bionic Flying Fish with Applications . . . 227
Haifeng Cai, Mei Liu, and Dan Su

Data Mining/Machine Learning

Decentralised and Privacy Preserving Machine Learning for Multiple Distributed Data Resources 235
Mona Alkhozae and Xiaojun Zeng

Markov Decision Processes with Contextual Nodes as a Method of Assessing Attacking Player Performance in Rugby League 251
Thomas Sawczuk, Anna Palczewska, and Ben Jones

Rule Base Structure and Influence of Existing Rules: An Initial Investigation 264
Changhong Jiang, Changjing Shang, and Qiang Shen

Using AI Approaches for Predicting Tomato Growth in Hydroponic Systems 277
Gadelhag Mohmed, Steven Grundy, Ahmad Lotfi, and Chungui Lu

Interpretable Model for Artefact Detection in Local Field Potentials via Feature Extraction and Decision Trees 288
Marcos I. Fabietti, Mufti Mahmud, Ahmad Lotfi, Alberto Averna, David Guggenmos, Randolph Nudo, and Michela Chiappalone

Comparative Analysis of K-Means and Traversal Optimisation Algorithms 300
David Ada Adama, Timilehin Yinka Olatunji, Salisu Wada Yahaya, and Ahmad Lotfi

Emojional: Emoji Embeddings 312
Elena Barry, Shoaib Jameel, and Haider Raza

Q-Routing Using Multiple QoS Metrics in SDN 325
Douglas Harewood-Gill, Trevor Martin, and Reza Nejabati

Creating a Hierarchical Fuzzy System to Assess Physical Activity Levels from Fitbit Data 337
F. A. Chaudhry, J. M. Garibaldi, and N. Qureshi

Image Analysis

Improved SLAM Method Based on Dynamic Objects Detecting with the Fusion of Vision and Lidar 347
Ren Zhang, Yitian Li, and Lu Lou

Image Fusion for Remote Sizing of Hot High Quality Steel Sections . . . 357
Y. Lin, P. Wang, R. Muroiwa, S. Pike, and L. Mihaylova

Detecting and Tracking the Moving Vehicles Based on Deep Learning 369
Jia Yan, Zhengxia Wang, Yitian Li, Zhen Chen, Xiaoyun Chen, and Lu Lou

Fast Lane Detection Based on Improved Enet for Driverless Cars 379
Boyi Li, Yi Zhao, and Lu Lou

Fast and Accurate 3D Reconstruction of Plants Using MVSNet and Multi-View Images 390
Zhen Chen, Hui Lv, Lu Lou, and John H. Doonan

An Improved SLAM Based on the Fusion of Lidar and Monocular Vision 400
Yitian Li, Ren Zhang, Qihuan Li, and Lu Lou

Segmentation and Counting of Plant Organs Using Deep Learning and Multi-view Images 406
Hui Lv, Zhen Chen, Yuhang Mo, Lu Lou, Ran Song, and John H. Doonan

Health Informatics

Estimation of Event-Related Potentials from Single-Trial EEG 415
Siyuan Zang, Changle Zhou, and Fei Chao

Proposing a Three-Stage Model to Quantify Bradykinesia on a Symptom Severity Level Using Deep Learning 428
R. Jaber, Rami Qahwaji, Amr Abdullatif, J. Buckley, and R. Abd-Alhameed

Developing Prediction Models for Large for Gestational Age Infants Using Ethnically Diverse Data 439
Sumaia Sabouni, Rami Qahwaji, Krzysztof Poterlowicz, and Anne M. Graham

Assessing Significance of Cognitive Assessments for Diagnosing Alzheimer's Disease with Fuzzy-Rough Feature Selection 450
Tianhua Chen, Changjing Shang, Pan Su, Yinghua Shen, Mufti Mahmud, Raymond Moodley, Grigoris Antoniou, and Qiang Shen

Towards Fuzzy Context-Aware Automatic Gait Assessments in Free-Living Environments 463
Megan Taylor Bunker, Arshad Sher, Vera Akpokodje, Federico Villagra, Neil Mac Parthaláin, and Otar Akanyeti

ASD Children Adaption Behaviour Assessment via Hand Movement Properties: A RoadMap 475
Dinghuang Zhang, Carrie M. Toptan, Shuwen Zhao, Gongyue Zhang, and Honghai Liu

Engineering Data- and Model-Driven Applications

A Model-Based Reliability Analysis Method Using Bayesian Network 483
Sohag Kabir and Felician Campean

Dealing with Uncertainty in Anomalous Audio Event Detection Using Fuzzy Modeling 496
Zied Mnasri, Stefano Rovetta, Francesco Masulli, and Alberto Cabri

A Framework to Handle Uncertainties of Machine Learning Models in Compliance with ISO 26262 508
Vinod Vasudevan, Amr Abdullatif, Sohag Kabir, and Felician Campean

Cloud-Agnostic and Lightweight Big Data Processing Platform in Multiple Clouds Using Docker Swarm and Terraform 519
Nitin Naik

Defect Prediction on Production Line 532
Souhaiel Khalfaoui, Eric Manouvrier, Alexandre Briot, David Delaux, Stéphane Butel, Jesutofunmi Ibrahim, Tatenda Kanyere, Bola Orimogunje, Amr Abdullatif, and Daniel Neagu

Driver Behaviour Modelling: Travel Prediction Using Probability Density Function 545
Alexey Uglanov, Kirill Kartashev, Felician Campean, Aleksandr Doikin, Amr Abdullatif, Emanuele Angiolini, Chunxing Lin, and Qichun Zhang

Driver Behaviour Clustering Using Discrete PDFs and Modified Markov Algorithm 557
Kirill Kartashev, Aleksandr Doikin, Felician Campean, Alexey Uglanov, Amr Abdullatif, Qichun Zhang, and Emanuele Angiolini

Towards Autonomous Health Monitoring of Rails Using a FEA-ANN Based Approach 569
Luke Brown, Shukri Afazov, and Daniele Scrimieri

Author Index 577

Fuzzy Systems

An Evolving Feature Weighting Framework for Granular Fuzzy Logic Models

Muhammad Zaiyad Muda(✉) and George Panoutsos

Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, UK
mzmuda1@sheffield.ac.uk

Abstract. Discovering and extracting knowledge from large databases are key elements in granular computing (GrC). The knowledge extracted, in the form of information granules can be used to build rule-based systems such as Fuzzy Logic inference systems. Algorithms for iterative data granulation in the literature treat all variables equally and neglects the difference in variable importance, as a potential mechanism to influence the data clustering process. In this paper, an iterative data granulation algorithm with feature weighting called W-GrC is proposed. By hypothesising that the variables or features used during the data granulation process can have different importance to how data granulation evolves, the weight of each feature's influence is estimated based on the information granules on a given instance; this is updated in each iteration. The feature weights are estimated based on the sum of within granule variances. The proposed method is validated through various UCI classification problems:- Iris, Wine and Glass datasets. Result shows that for certain range of feature weight parameter, the new algorithm outperforms the conventional iterative granulation in terms of classification accuracy. We also give attention to the interpretability-accuracy trade-off in Fuzzy Logic-based systems and we show that W-GrC produces higher classification performance - without significant deterioration in terms of its interpretability (Nauck's index).

Keywords: Granular computing · Iterative data granulation · Fuzzy logic · Feature weights · Feature relevance

1 Introduction

One of the key steps in building data driven Fuzzy Logic (FL) models is the process of extracting knowledge from data [1]. Granular Computing (GrC) and iterative data granulation algorithms are an effective approach to extract knowledge from data within the context of human-centric systems [2, 3]. Among the most widely used techniques for this process are fuzzy c-means (FCM) and hierarchical clustering.

In recent years, iterative data granulation algorithms, also known as granular clustering proposed in [3, 4] have become a proven alternative in data mining and developing FL rule-bases. The main idea of this algorithm is to merge two most compatible information granules iteratively until sufficient data compression is achieved [3]. The

T. Jansen et al. (Eds.): UKCI 2021, AISC 1409, pp. 3–14, 2022.
https://doi.org/10.1007/978-3-030-87094-2_1

compatibility measure can simply be distance based (such as in hierarchical clustering algorithms) or potentially involve more complex formulations that combine granular density, cardinality, overlap etc.

GrC algorithms have a similarity with agglomerative hierarchical clustering in terms of its 'find and merge' strategy. However, one main distinction between these algorithms is that in GrC, every granule consist of sub-granules originating directly from the actual data. This contributes to strong connection between the raw data and the information granules. Moreover, the compatibility measure in GrC is very useful tool in monitoring the similarity between granules; this can be linked to a numerical criterion to terminate the granulation in order to avoid merging of low compatibility granules [4].

So far GrC algorithms treat all features equally during the data granulation process. This is not desirable especially when dealing with data consisting of high number of features [5]. In many cases, some features are not as crucial as other features in the development of the FL model [1], while other features may have an importance that changes throughout the granulation process. This leads to the idea of continuously measuring and assigning appropriate weight for each feature throughout the data granulation (as in adaptive feature weighting for classical clustering algorithms).

Even though the feature weight concept has been introduced elsewhere [4], most of the works regarding this algorithm such as [6] and [7] use fixed weight for each feature. Investigations in feature weighting for GrC are scarce; for example in [8] a Fast Correlation-Based Filter which is based on symmetrical uncertainty to determine the most relevant features of a welding process. However, this is a preprocessing step (acting as a filter method) where the feature weights are determined in advance, and their values are constant throughout the evolution of the granulation process.

In this paper, we propose a new GrC algorithm that assigns and updates the feature weights based on the importance of the input features throughout the evolution of the iterative data granulation. With this approach we enable the more important features to have higher influence in the data granulation than the less important features, for a given iteration. Furthermore, instead of assigning the weight in the preprocessing step, the feature weighting in this research is embedded in the granulation process itself. This allows the feature weights to be adjusted according to the information granules that have been formed. The hypothesis here is that as information granules merge, and patterns develop, the importance of particular features to the evolution of such granules may change. While this approach is new in GrC, it has already been proven to be effective in other data mining and clustering algorithms.

Feature weighting has been applied in many clustering algorithms to overcome the problem of feature selection. Wu et al. introduced a new weighted fuzzy c-means algorithm taking into account the between-cluster separability [9]. The iterative formulas of the feature weights and membership degrees are obtained by maximizing the in-cluster compactness and the between-cluster separability. In another research, Huang et al. proposed W-*k*-means, the weighted version of k-means that outperformed the standard k-means in recovering clusters in data [5]. They also demonstrated that eliminating the irrelevant features based on the feature weights may enhance the clustering results. In the area of hierarchical clustering, Amorim implemented the feature-weighting scheme in an improved version of Ward, called Ward_p [10]. He demonstrated the effectiveness of Ward_p over the conventional Ward in particular in datasets comprising noisy data.

2 Background: The GrC-Fuzzy Logic Model

The general framework for GrC-Fuzzy Logic modelling consists of two main steps, which are knowledge discovery and followed by the formation of a Fuzzy Logic rule-base. In the knowledge discovery step, granular computing and the process of iterative data granulation mimic the cognitive human abstraction in grouping entities with similar features (i.e. geometrical properties, cardinality, density, etc.) [6]. The knowledge discovered in the form of information granules defines the structure of the FL rule-base, specifically the parameters of the FL membership functions.

2.1 Knowledge Discovery

The process of iterative data granulation starts with finding the pair of granules with the highest compatibility measure. Next, the granules are merged together in a new information granule that consists of original granules [7]. These steps are repeated until a satisfactory data abstraction level is accomplished.

The compatibility measure of two granules A and B is defined as:

$$C(A, B) = Distance_{MAX} - Distance_{A,B}.\exp(-\alpha \times R) \tag{1}$$

where

$$Density\ R = \frac{C_{A,B}/Cardinality_{MAX}}{L_{A,B}/Length_{MAX}} \tag{2}$$

$Distance_{MAX}$ is defined as the sum of maximum distance in each dimension d:

$$Distance_{MAX} = \sum\nolimits_{v=1}^{d}(distance_v) \tag{3}$$

$Distance_{A,B}$ is the average multidimensional distance between granules A and B weighted by feature weight w_v:

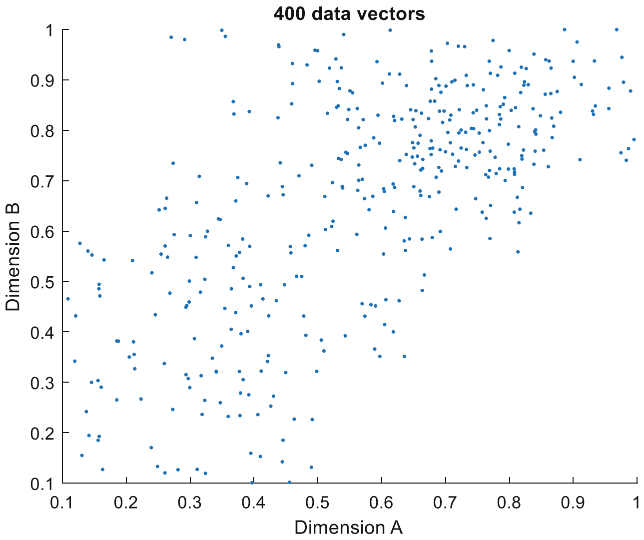
$$Distance_{A,B} = \frac{\sum_{v=1}^{d} w_v(D_1 - D_2)}{\text{d}} \tag{4}$$

in which

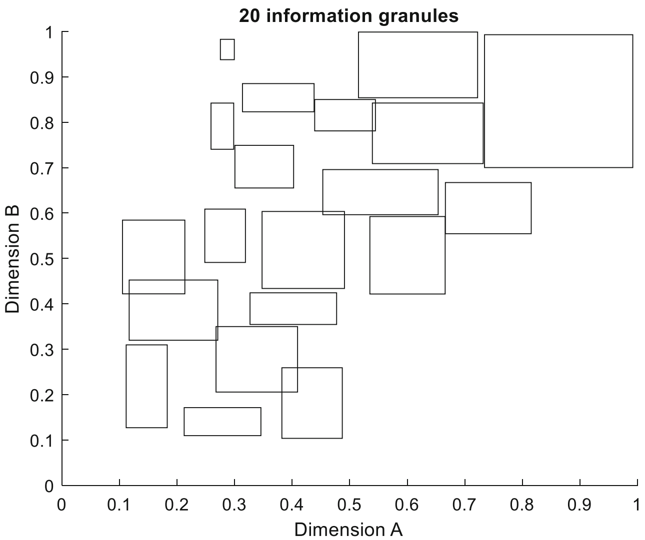
$$D_1 = \max(max_{Av}, max_{Bv}) \tag{5}$$

$$D_2 = \min(min_{Av}, min_{Bv}) \tag{6}$$

max_{Av}: maximum value in granule A for dimension v, min_{Av}: minimum value in granule A for dimension v, α: weights the requirement between distance and density, $Cardinality_{MAX}$: the total number of instances in the data set, $Length_{MAX}$: the maximum

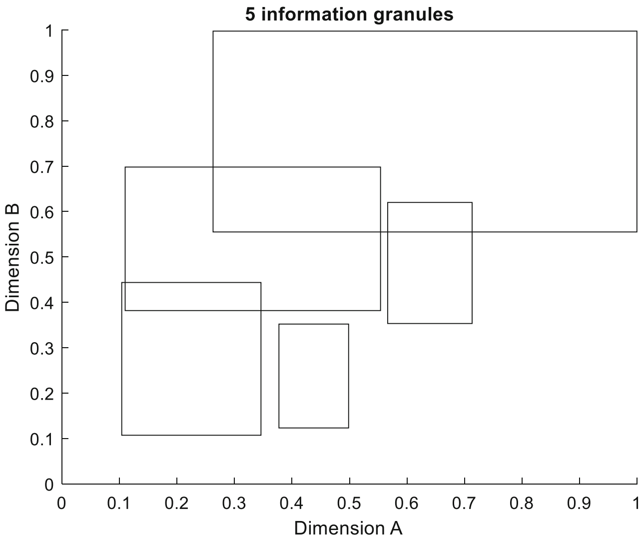


(a)



(b)

Fig. 1. Data granulation process from (a) 400 data vectors to (b) 20 information granules and (c) 5 information granules



(c)

Fig. 1. continued

possible length of a granule in the data set, $C_{A,B}$: the cardinality when granule A merge with granule B, and $L_{A,B}$: length of the granule A and B, defined as:

$$L_{A,B} = \sum\nolimits_{i=v}^{d} (max_{X_v} - min_{X_v}) \tag{7}$$

Typically, the feature weight parameter w_v in Eq. (4) used in most previous works is set to 1 for all dimensions (i.e. feature weighting is not used), or used at a fixed pre-determined value for each feature. The computation and adaptive adjustment of this parameter is the focus in this paper.

Figure 1 illustrates the evolution of a data granulation process for a two-dimensional synthetic data set with 150 instances. It starts with the initial raw data where every data instance is considered as one granule-point. These granules are then merged iteratively causing the number of granules to be reduced until the final information granules are established (in a predetermined manner, or using some termination criterion).

2.2 Formation of Fuzzy Logic Rule-Base

Taking a Gaussian Fuzzy Logic membership function (MF) as an example, the MF depends on two parameters σ and c which represent the width and the centre of a fuzzy set [11]. The standard deviation and median of data in each information granule can be used to determine the σ and c, respectively. Each information granule characterises one fuzzy rule [12]. For example, five information granules in Fig. 1 will lead to the formation of five fuzzy rules. Figure 2 shows the overview of GrC-Fuzzy Logic modelling framework.

By determining the parameters σ and c across each input dimension individually in a multi-input single-output (MISO) system, the rules based on Mamdani fuzzy inference system (FIS) can be written as follows:

$$\begin{gathered} Rule1: \\ IF(inputA = A_1 \ and \ inputB = B_1 \ and \ \ldots)then(output = O_1) \\ Rule2: \\ IF(inputA = A_2 \ and \ inputB = B_2 \ and \ \ldots)then(output = O_2) \end{gathered} \tag{8}$$

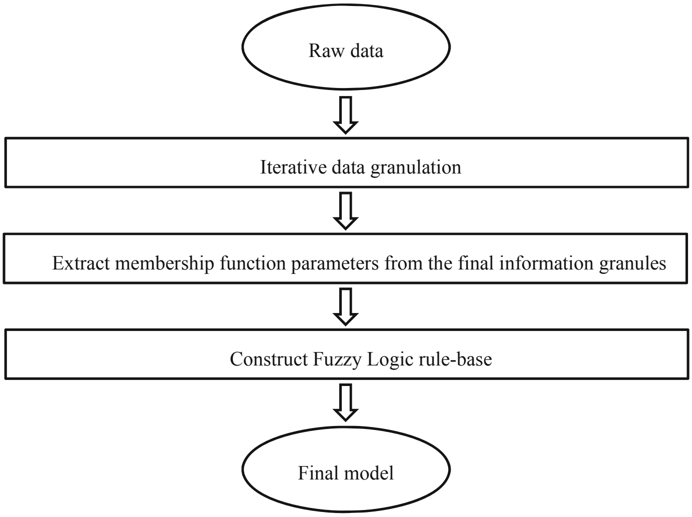


Fig. 2. The overview of GrC-Fuzzy Logic modelling framework

3 Proposed Methodology: Evolving Feature Weighting GrC

The Weighted K-Means (WK-Means) algorithm introduced by Huang et al. [5] minimises the following object function:

$$W(S, C, w) = \sum_{k=1}^{K} \sum_{i \in S_k} \sum_{v \in V} w_v^{\beta} d(y_{iv}, c_{kv}) \tag{9}$$

The Equation above is minimised by an iterative method, optimising (9) for S, C, and w, where $S = \{S_1, S_2, \ldots, S_k, \ldots, S_K\}$, $c_k \in C$ is the centroid for each granule k, y_i is an object in dataset Y, and β is the feature weighting parameter that balances the degree of effect between the weight and its contribution to the distance. There are two possibilities for the update of w_v, with S and C fixed, subject to $\beta > 1$:

$$w_v = \begin{cases} 0, & if \ D_v = 0 \\ \dfrac{1}{\sum_{j=1}^{h} \left[\frac{D_v}{D_j}\right]^{\frac{1}{\beta-1}}}, & if \ D_v \neq 0 \end{cases} \tag{10}$$

where h is the number of features where $D_v \neq 0$.

The parameter w_v (feature weight) in Eq. (4) has a fixed value, often pre-determined, in works related to GrC. In this paper, the weight for each feature v is defined and iteratively updated based on Eq. (10).

As shown in the equation, nonzero weight is only assigned to a feature where $D_v \neq 0$. $D_v = 0$ indicates that the *vth* feature consists of single value in each granule [5] and will be assigned zero weight. In this research, D_v is set as the sum of within granule variance:

$$D_v = \sum\nolimits_{k=1}^{K} \frac{1}{N-1} \sum\nolimits_{i=1}^{N} |y_{iv} - c_{kv}|^2 \tag{11}$$

where N is the cardinality in the granule k.

The underlying principle here is to assign higher weights for features with lower within granule variance i.e. high variance in granules is set to be undesirable, hence penalised in the compatibility index. High variance would translate into high *sigma* (width) MFs. Hence, features that drive the creation of low variance granules, in any given iteration step, are promoted by the use of this adaptive weight and such features are considered here as more important for the evolution of the granulation process towards the development of FL rule-bases for classification problems.

3.1 Simulations and Empirical Results

Simulations were conducted on three datasets with regard to classification problems:– Iris, Wine and Glass (UCI Machine Learning Repository). All features are scaled to the interval of [0, 1]. The ratio of training and testing data is set to 80:20. The range of feature weighting parameter β is selected between 1.5 and 10. The root mean square error (RMSE) and prediction accuracy % were calculated as the average of ten trials.

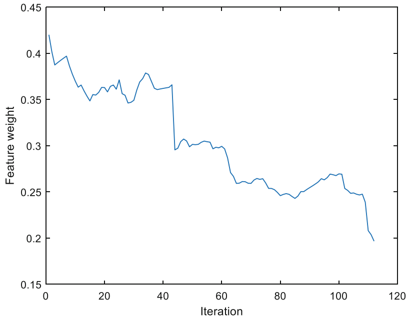
The Iris data consists of 150 instances with four input features. Next, the experiment is scaled up to datasets with higher feature dimensionality, which are Glass and Wine data with 10 and 13 input features, respectively. A bootstrapping method is applied to Glass data to balance the number of instances for each class. Due to this, the number of instances increases from 214 to 371. For comparison purposes, based on previous work [12], the number of granules selected for Iris and Wine is 5, while for Glass is 30 granules.

3.2 Evolving Feature Weights

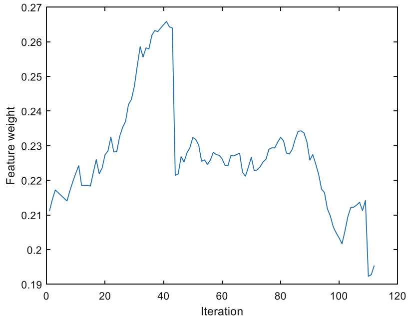
Figure 3 shows how the feature weights evolve throughout the iterative granulation process, as an example for two features in the Iris dataset. The weights are plotted starting from the fourth iteration (out of 115 iterations), after which the feature weights are observed to be stable. This is due to the fact that the feature weights are assigned based on the within granule variance, while the merging process at the beginning only involves singleton granules (i.e. $D_v = 0$).

The feature weight average is computed and is compared with other measures such as mutual information and feature importance score as shown in Table 1. Mutual information gives information about the relevance between two random variables and normally being estimated between each feature and the given class labels [13]. The feature importance score ranks the features using a chi-square (χ^2) test [14]. The feature importance score is the negative log of chi-square tests' p-value [15].

This result shows that the feature weight ranking is consistent with the other two independent measures, confirming our hypothesis in capturing feature importance via the proposed method. All these three measures rank Petal width as the most important feature, followed by Petal length, Sepal length and Sepal width.



(a)



(b)

Fig. 3. Feature weights throughout the granulation for (a) Sepal length and (b) Sepal width

Table 1. Comparison of average feature weight in W-GrC with the feature importance score and mutual information

	Average weight (W-GrC)	Feature importance score	Mutual information
Sepal length	0.2721	41.7358	0.6415
Sepal width	0.2062	19.1551	0.3935
Petal length	0.3072	97.8866	1.2663
Petal width	0.3623	101.1028	1.3245

3.3 Empirical Results Using Simulations

Table 2 summarises the performance of W-GrC with different values of β. The 'no feature weighting' row presents the results for the GrC without feature weighting, also known as conventional GrC. It is observed that with careful selection of β, the proposed W-GrC outperforms the standard GrC in terms of RMSE and accuracy. β needs to be treated as a hypermeter here, which will be identified in each case (problem specific).

For the Iris data, good results were obtained at $\beta \in \{3, 4, 5, 6, 7, 8, 10\}$. The highest accuracy was achieved when $\beta = 3$ and $\beta = 6$ with 96.33% of correct prediction as compared to 94% in the conventional GrC. For Wine data, improvement can be observed at β ranging from 3 to 6. Most experiments showed accuracies of above 90%, except for

$\beta = 1.5$. This result is comparable to other literature results, however it is recognised that this specific case study may be too simple to stress test the proposed methodology (Glass and Wine data offer higher complexity and dimensionality).

In the case of the Glass dataset, we can see more clearly that higher values of β ($\beta \geq 3$) are more desirable to produce good result. The best performance is recorded at $\beta = 5$ with 71.86% accuracy in comparison with 62.79% in conventional GrC.

From Table 2, it can be observed that in general, W-GrC outperforms the conventional GrC. It achieves highest accuracy for all datasets, when an appropriate value of β is selected. This is because features that are more important for a given instance during the iterative granulation process are assigned with larger weights in forming the information granules. However, it is noted that the selection of β is important to obtain high classification accuracy. From the result, we suggest ($\beta \geq 3$) as the appropriate value of β, for this particular case study.

Results are benchmarked against other research such as [16] with 96.67% in Iris, [17] with 97.14% in Wine and [13] with 71.66% in Glass.

Table 2. Average RMSE and % accuracy performance of W-GrC with various β values, testing (unseen) data, 10 runs per β value

	Iris		Wine		Glass	
	RMSE	Accuracy (%)	RMSE	Accuracy (%)	RMSE	Accuracy (%)
No feature weighting	0.1415	94	0.1173	92.3	0.2020	62.79
$\beta = 1.5$	0.1473	91.67	0.3101	66.67	0.4274	26.74
$\beta = 2.0$	0.1551	90.67	0.1238	91	0.3365	32.33
$\beta = 3.0$	0.1205	96.33	0.1082	94	0.2235	63.02
$\beta = 4.0$	0.1302	94.67	0.1123	92.67	0.2164	69.30
$\beta = 5.0$	0.1253	94.33	0.1033	95.67	0.2144	71.86
$\beta = 6.0$	0.1251	96.33	0.1067	93	0.2165	66.98
$\beta = 7.0$	0.1285	95.67	0.1230	92	0.1980	66.51
$\beta = 8.0$	0.1189	96	0.1342	90.33	0.2219	66.05
$\beta = 9.0$	0.1346	93.67	0.1186	91.67	0.2105	68.14
$\beta = 10.0$	0.1273	95	0.1212	91.33	0.2224	65.81

3.4 Interpretability Index

In designing Fuzzy Logic systems (FLS), interpretability and accuracy are often conflicting objectives; one can be enhanced by sacrificing the other, a situation that is termed as interpretability-accuracy trade-off. For example the enhanced interpretability of Mamdani-based FLS, versus the enhanced predictive accuracy of TSK-based FLS. Interpretability, within the FLS context, can be defined as the trait of a model to enable human to understand a system's behavior by scrutinising its rule base [18]. In this study, we use the models developed using values of β that perform the best in terms of accuracy as in Table 2 to assess if the models' interpretability is affected by the enhanced predictive performance.

The impact of feature weighting on interpretability measure is investigated using Nauck's index. Nauck's index is a numerical index introduced by Nauck in order to assess the interpretability of FL rule-based classification systems [19, 20]. It is computed as the product of three terms: complexity of FLS (*comp*), average normalized coverage of fuzzy partition ($\overline{cov}$) and average normalized partition index ($\overline{part}$) given by:

$$Nauck\ index = comp \times \overline{cov} \times \overline{part} \tag{12}$$

(readers are referred to [19] and [20] for further details).

Table 3 summarises the comparison of the interpretability index for the proposed W-GrC and the conventional GrC. It is demonstrated that W-GrC is able to producing higher accuracy without a statistically significant deterioration in terms of model interpretability. The impact on interpretability index is minor, less than 2% on the Iris data, and even less for the Wine and Glass case studies. Note that the Nauck's index in Glass is comparatively to the other cases small due to the high number of rules (30 as opposed to 5 in Iris and Wine).

Table 3. Comparison of the interpretability index

	Nauck's index	
	W-GrC	GrC
Iris	0.3076	0.3129
Wine	0.0929	0.0928
Glass	7.02×10^{-4}	7.07×10^{-4}

4 Conclusion

In this paper, a new iterative data granulation algorithm is presented with evolving feature weighting to characterise the importance of data features and use such weights to drive the information granulation process. The weight for each feature is determined based on the sum of within granule variances from the granules that have been formed, at any given iteration. In each iteration, the importance of all features is evaluated to identify the most important features that contribute most to the computation of the granules' compatibility measure.

The resulting importance of features, estimated via averaging feature weights throughout the data granulation process, are compared with other methods such as chi-square test and mutual information; agreement in feature ranking is demonstrated.

Simulation results in UCI classification problems have shown that the proposed W-GrC algorithm outperforms the conventional GrC in terms of classification accuracy. Improvement can be seen, in more complex datasets such as Glass case study. The experiment results showed that the proposed GrC-Fuzzy-modelling framework is able to handle data with various dimensionality. The interpretability of the resulting models is assessed, using Nauck's index, and no significant deterioration of predictive performance is observed despite the higher resulting % accuracy in the classification tasks. While this study shows positive preliminary results, a greater range of complexity in case studies can be investigated in the future, as well as performance can be assessed more extensively against non GrC-based methods.

Acknowledgement. This research is sponsored by Universiti Teknologi MARA, Ministry of Education, Malaysia and The University of Sheffield, UK.

References

1. Ahmad, S.S.S., Pedrycz, W.: Data and feature reduction in fuzzy modeling through Particle Swarm Optimization. Appl. Comput. Intell. Soft Comput. (2012)
2. Hu, J., Li, T., Wang, H., Fujita, H.: Hierarchical cluster ensemble model based on knowledge granulation. Knowl.-Based Syst. **91**, 179–188 (2016)
3. Pedrycz, W., Bargiela, A.: Granular clustering: a granular signature of data. IEEE Trans. Syst. Man Cybern. Part B: Cybern. **32**(2), 212–224 (2002)
4. Panoutsos, G., Mahfouf, M.: A neural-fuzzy modelling framework based on granular computing: concepts and applications. Fuzzy Sets Syst. **161**(21), 2808–2830 (2010)
5. Huang, J.Z., Ng, M.K., Rong, H., Li, Z.: Automated variable weighting in k-means type clustering. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 657–668 (2005)
6. Rubio-Solis, A., Panoutsos, G.: Granular computing neural-fuzzy modelling: a neutrosophic approach. Appl. Soft Comput. **13**(9), 4010–4021 (2013)
7. Baraka, A., Panoutsos, G., Mahfouf, M., Cater, S.: A Shannon entropy-based conflict measure for enhancing granular computing-based information processing. In: 2014 IEEE International Conference on Granular Computing (GrC), pp. 13–18 (2014)
8. Rubio-Solis, A., Panoutsos, G.: An ensemble data-driven fuzzy network for laser welding quality prediction. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6 (2017)
9. Wu, Z., Li, C., Zhou, F., Liu, L.: A new weighted fuzzy c-means clustering approach considering between-cluster separability. J. Intell. Fuzzy Syst. **40**(1), 1017–1024 (2021)
10. Amorim, R.: Feature relevance in Ward's hierarchical clustering using the Lp norm. J. Classif. **32**, 46–62 (2015)
11. Iliadis, L., Skopianos, S., Tachos, S., Spartalis, S.: A fuzzy inference system using Gaussian distribution curves for forest fire risk estimation. In: Papadopoulos, H., Andreou, A.S., Bramer, M. (eds.) AIAI 2010. IAICT, vol. 339, pp. 376–386. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-16239-8_49
12. Muda, M.Z., Panoutsos, G.: An entropy-based uncertainty measure for developing granular models. In: 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI), pp. 73–77 (2020)
13. Xing, H., Ha, M., Tian, D., Hu, B.: A novel support vector machine with its features weighted by mutual information. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), pp. 315–320 (2008)

14. Alkhodari, M., Islayem, D.K., Alskafi, F.A., Khandoker, A.H.: Predicting hypertensive patients with higher risk of developing vascular events using heart rate variability and machine learning. IEEE Access **8**, 192727–192739 (2020)
15. Fernando, H., Marshall, J.: What lies beneath: material classification for autonomous excavators using proprioceptive force sensing and machine learning. Autom. Constr. **119** (2020)
16. Wang, Y., Cang, S., Yu, H.: Mutual information inspired feature selection using kernel canonical correlation analysis. Expert Syst. Appl. **X 4** (2019)
17. Dousthagh, M., Nazari, M., Mosavi, A., Shamshirband, S., Chronopoulos, A.T.: Feature weighting using a clustering approach. Int. J. Model. Optim. **9**(2), 67–71 (2019)
18. Shukla, P., Tripathi, S.: A review on the interpretability-accuracy trade-off in evolutionary multi-objective fuzzy systems (EMOFS). Information **3**, 256–277 (2012)
19. Razak, T.R., Garibaldi, J., Wagner, C., Pourabdollah, A., Soria, D.: Interpretability indices for hierarchical fuzzy systems. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2017) (2017)
20. Nauck, D.D.: Measuring interpretability in rule-based classification systems. In: The 12th IEEE International Conference on Fuzzy Systems, vol. 1, pp. 196–201 (2003)

Fuzzy Multi-Criteria Decision-Making: Example of an Explainable Classification Framework

Hesham Yusuf[1,2(✉)], Kai Yang[2], and George Panoutsos[1]

[1] University of Sheffield, Western Bank, Sheffield S10 2TN, UK
hahyusuf1@sheffield.ac.uk
[2] TWI Ltd., Great Abington, Cambridge CB21 6AL, UK

Abstract. Explanation, or system interpretability, has always been important in applications where critical decisions need to be made, for example in the justice system or biomedical applications. In artificial intelligence and machine learning, there is an ever increasing need for system interpretability. This paper investigates a Fuzzy Multi-Criteria Decision-Making (MCDM) model as the basis for an interpretable framework for explainable classification. The proposed framework includes a Fuzzy Inference System paired with a modified MCDM-based model for data-driven classification. The modular nature of MCDM allows for the development of a model-based layer capable of generating factual and counterfactual explanations. Results on a 'Titanic' survivors' dataset classification, which illustrates a minimal trade-off in predictive performance while gaining textual and graphical explanation, autonomously provided by the proposed model-based MCDM framework.

Keywords: Fuzzy logic · Interpretability · Multi-criteria decision making · Explainable-AI

1 Introduction

Interpretability has been a topic of significant interest among researchers with the vision that it could shape how machine learning (ML) frameworks are adopted in the future [1–4]. The current state-of-the-art ML classification frameworks are not necessarily interpretable; a property of models that could enable *explainability* of the models' results. With the advent of deep learning and the power of high performance computing, data-driven ML seems to be the obvious choice for data-rich tasks. For the most part, deep learning and other state-of-the-art ML techniques are sufficiently accurate predictive models and, are constructed with data paired with minimal, if any, expert knowledge. The challenge with such models is the fact they are often black box models; hence, they are neither inherently interpretable nor explainable. The lack of transparency is an obstacle to the wide adoption of such methods, especially in applications requiring precise decision justification [1,4,5]; for example, safety critical applications such as nuclear, medical and advanced manufacturing.

T. Jansen et al. (Eds.): UKCI 2021, AISC 1409, pp. 15–26, 2022.
https://doi.org/10.1007/978-3-030-87094-2_2

The focus of this paper, is multi-criteria decision making, and in particular interpretable data-driven Fuzzy-Multi-Criteria Decision-Making (Fuzzy-MCDM) for classification problems. In this Section, a literature review summary on MCDM, interpretability and explainability are covered. The methodology used, in Sect. 2, is an expansion of Fuzzy-Amended fused TOPSIS-VIKOR for Classification (Fuzzy-ATOVIC) [11], a MCDM-based technique developed for achieving satisfactory performance while being interpretable. Fuzzy-ATOVIC is consequently augmented with an explanation framework designed for explaining the classification output. Section 3 includes the framework's application to the Kaggle 'Titanic' dataset, which presents a classification problem [6]. The results demonstrate the model's ability to provide graphical and textual explanation, while maintaining comparable classification performance. The paper finishes with concluding remarks and future work.

Multi-Criteria Decision Making (MCDM) is a set of modelling methods capable of providing decision support based on several criteria [7]. MCDM are applied in a variety of applications such as business, supply chain and manufacturing [8]. The methods often use a range of criteria to determine a *rank* for each *object*. An example of a typical MCDM application is the ranking of a supplier list. In this case, a company would compare a set of suppliers by using a set of criteria such as speed of delivery, pricing, and payment terms. Depending on the circumstances, different levels of importance can be assigned to different criteria using weights. The process results in a ranked list with the alternatives. Although MCDM was not initially intended as a classifier, nevertheless, there were attempts of developing MCDM-based classification frameworks [9,10]. MCDM can utilise human-understandable criteria, hence it can become interpretable by nature deeming it a viable candidate for explainable AI systems, when combined with AI-based methods.

Amended fused TOPSIS-VIKOR for Classification (ATOVIC) is a supervised learning MCDM framework that can be trained by a combination of data and expert knowledge [9]. A Fuzzy Logic-based extension of the method, Fuzzy-ATOVIC, makes use of a Fuzzy Inference System (FIS) to replace the final step in the decision making process, introducing greater potential for interpretability to the overall MCDM framework [11]. Fuzzy-ATOVIC as an initial proposal was the first step towards adapting ATOVIC as a fully data-driven classification framework while maintaining its interpretability [11]. Achieving explainability in complex ML structures has always been a challenge due to the inherent non-interpretable nature of many such models. The lack of explanatory information in such models delayed the long awaited wide adoption in several industries. Explanation, as a functional requirement, is considered important in areas where the wrong decision is likely to have a major or catastrophic consequence. In these applications, it is imperative that ML models can provide *explanation* because without it, the user is faced with relying on their own calculations to make decisions, defying the ultimate purpose of the model - improving the overall efficiency and accuracy of the process.

Interpretability has two main categories: model-based or post-hoc [3]. Model-based, as the name suggests, is interpretability that utilises the model itself (its parameters and variables), as the source of interpretation. In contrast, post-hoc interpretability relies solely on the input(s) and output(s) as the source of interpretation [3]. Many researchers have attempted to utilise post-hoc to attempt to explain the output [12,13]. One of the weaknesses of post-hoc interpretability is the fact that it does not directly explain how the model arrived at a certain decision, rather it is in some way an explanation estimator. On that account, model-based interpretability offers the potential for a direct explanation of the models' decision making process. One of the main challenges in pursuing model-based interpretability is to overcome the trade-off of performance (accuracy, interpretability) [1].

2 Methodology

2.1 ATOVIC and Fuzzy-ATOVIC

Amended Fused TOPSIS-VIKOR for Classification (ATOVIC) is an MCDM-based classification technique introduced by Baccour in 2018 [9]. ATOVIC is a fusion of two MCDM-based techniques: TOPSIS and VIKOR [9]. As opposed to most MCDM techniques, ATOVIC is supervised and data-driven: however, it relies on expert knowledge for setting whether a feature is a *cost* or *benefit*. It is vital to set features as costs or benefits effectively to maximise performance. Furthermore, relying on expert knowledge for data-driven applications could be problematic for datasets where such knowledge does not exist; thus, a method was implemented, as will be explained in Step 2 of model construction, to numerically classify a feature as a cost or benefit. Fuzzy-ATOVIC is a fuzzy extension of ATOVIC that uses a Fuzzy Inference System (FIS) for the final step of decision making [11].

Construction of the ATOVIC model is achieved using the following steps. The procedure is based on Baccour's literature [9], while steps 2 and 3 were modified to enhance the methods of weight calculation and feature classification; to improve the accuracy performance and eliminate the requirement of expert knowledge.

1. Training dataset normalisation using (1, 2) where θ is the normalised term, r denotes the reference matrix, x is the non-normalised term and h is the normalisation coefficient, p is the class number, i is the object number and j is the feature number.
2. Weight calculation using (3) where w_j is the weight and ρ_j is the Pearson correlation coefficient [15]; of feature j.
3. Classifying features as a benefit or cost is determined using ρ. If $\rho_j \leq 0$ then j is a cost for Class 2 and a benefit for Class 1; while if $\rho_j > 0$ then j is a cost for Class 1 and a benefit for Class 2. Where j is the feature number. To achieve this, the labels for class 1 and 2 data have to be set as 1 and 2 respectively.
4. Ideal solutions calculation using (4, 5) where two sets of ideal solutions f are calculated: positive and negative. For positive ideal solutions f_p^+, the maximum is used for a *benefit* feature while the minimum is used for a *cost*

feature. Intuitively, it is vice versa for negative ideal solutions, as shown in (5). The ideal solutions are later used for classification.

$$\theta^r_{ij_p} = \frac{x^r_{ij}}{h^r_{j_p}} \tag{1}$$

$$h_{j_p} = \sqrt{\sum_{i=1}^{m^r} (x^r_{ij_p})^2} \tag{2}$$

$$w_j = \frac{\rho^r_j}{\sum_{j=1}^{n} \rho^r_j} \tag{3}$$

$$\begin{aligned} f_p^+ &= \{\theta_1^{r^+}, \theta_2^{r^+}, \ldots, \theta_n^{r^+}\} \\ &= \Big\{(max_i \theta^r_{ij_p}/j \in B), (min_i \theta^r_{ij_p}/j \in C)\Big\} \end{aligned} \tag{4}$$

$$\begin{aligned} f_p^- &= \{\theta_1^{r^-}, \theta_2^{r^-}, \ldots, \theta_n^{r^-}\} \\ &= \Big\{(min_i \theta^r_{ij_p}/j \in B), (max_i \theta^r_{ij_p}/j \in C)\Big\} \end{aligned} \tag{5}$$

After model construction, the data is classified by executing the steps below.

1. Testing data normalisation using (1) and, based on the values of $h^r_{j_p}$ defined during model construction.
2. Distance measures S and R are the Manhattan and Chebyshev distances, respectively. They are obtained by calculating the distance types from the ideal solutions for class c = 1 to 2. This implementation of ATOVIC, as opposed to the original version, does not use the Q measure - a weighted sum of S and R. The purpose is to improve traceability and simulatability [1].
3. Comparing distance measures for classification by use of a FIS.

$$S_{c_i} = \sum_{j=1}^{n} w_j * (f^+_{ij_c} - \theta^t_{ij_c})/(f^+_{ij_c} - f^-_{ij_c}),\ S_{c_i} \in [0, 1] \tag{6}$$

$$R_{c_i} = \max_j \Big[w_j * (f^+_{ij_c} - \theta^t_{ij_c})/(f^+_{ij_c} - f^-_{ij_c})\Big],\ R_{c_i} \in [0, 1] \tag{7}$$

The measures S and R (6, 7), are input into a FIS to compute the fuzzy class. The FIS has six inputs as defined by (8, 9) where ΔM_c is calculated for $M = \{S, R\}$ and class $c = 1$ to 2.

$$\Delta M_c = M_{c,2} - M_{c,1} \tag{8}$$

$$n_M = |\Delta M_2| - |\Delta M_1| \tag{9}$$

Furthermore, the input-output membership functions (MFs) were configured as below.

– S_c, R_c: two MFs: `class_1`, `class_2`
– n_S, n_R: two MFs: `positive` (positive outcome model is used for decision), `negative` (negative outcome model is used for decision)
– Output: four MFs: `class_c_strong`, `class_c_normal`, for class `c` = 1 to 2.

Consequently, a set of 16 rules were configured to cover all possible combinations of inputs; this includes cases where the two sub-models are in agreement or conflict. The S measures are utilised to take a decision, while the R measures translates to a higher chance of certainty; if it is in agreement with S. If the measures are in agreement, a *strong* output MF, corresponding to the class, is set while a *normal* one is used in the case of conflict, as illustrated in Fig. 1. The updated configuration of ATOVIC means the FIS had to be modified to process the measures S and R, instead of just the weighted sum measure Q; in the first iteration of Fuzzy-ATOVIC [11].

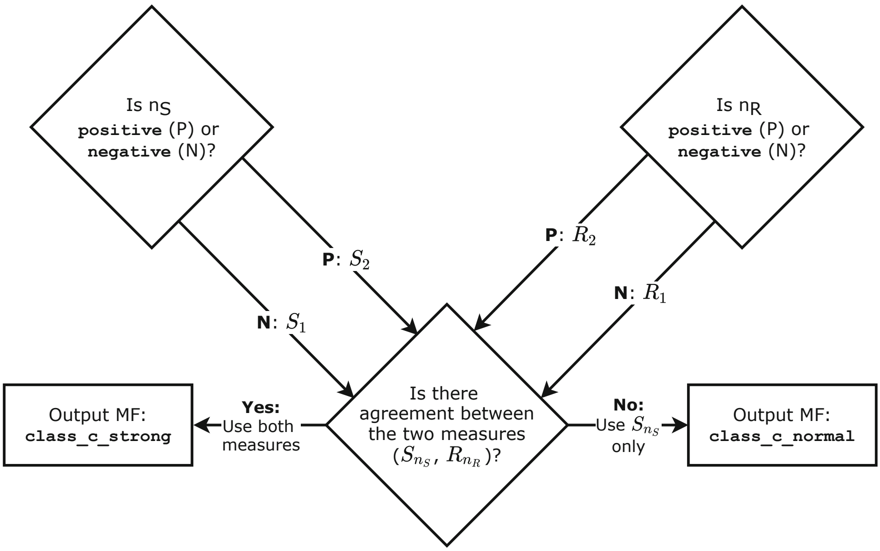


Fig. 1. Flowchart explaining the conditions for formulating the fuzzy rules: binary classification

Despite ATOVIC not utilising user-defined linguistic terms, using human-understandable features meant this is not necessary for interpretation. However, for features that are not human-understandable, it would be essential to introduce interpretability by pre-processing techniques suitable for the problem.