Granular Fuzzy Rule-Based Models: Design and Analysis

by

Xingchen Hu

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Software Engineering and Intelligent Systems

Department of Electrical and Computer Engineering
University of Alberta

Abstract

Fuzzy rule-based models have been studied for decades and emerged in a diversity of architectures and design approaches. They play a vital and unique role by forming a human-centric computing framework. In general, fuzzy rule-based models are regarded as numeric constructs; as such they are optimized and evaluated at the numeric level. However, no ideal fuzzy models that fully capture (coincide with) all numeric experimental data. Information granularity is a perspective to represent and recognize the abstraction of information as the observation method of humans. In this perspective, the numeric data can be presented at various levels of resolution or scales. By bringing a concept of information granularity into fuzzy rule-based models, we give up on obtaining precise numeric models, whereas we make them into granular form and produce granular results. Subsequently, the outputs provided by granular fuzzy rule-based models are aligned well with the experimental data and deliver better insight into credibility.

The fundamental objective of this thesis is to establish a comprehensive, systematic method for developing granular fuzzy rule-based models, so that the granular outputs of the models can embrace (cover) the target experimental data as much as possible, meanwhile the granular outputs are as specific as possible. To accomplish these objectives, we study several fundamental design issues that emerge in the realm of Granular Computing. First, we propose an advanced scheme of granulation and degranulation to

abstract information granules from numeric data. Second, we investigate several commonly known logic operators that are used in fuzzy modeling and granular fuzzy modeling. Afterwards, we design a series of development strategies for granular fuzzy rule-based models by admitting and allocating a certain level of information granularity around numeric values. Our proposed granular rule-based models could be classified into three groups: granular input space of the models, granular processing modules of the models, and granular output space of the models. Unlike the standard numeric-performance measure of fuzzy models that come in the form of the root-mean-square error (RMSE), two pertinent performance measures are introduced and implemented to evaluate the performance of granular fuzzy rule-based models: namely, coverage and specificity. We develop several protocols of forming and allocating information granules to cope with different strategies of granular modeling and analyze how different protocols lead to improve the performance of granular models. Some commonly used population-based optimization algorithms—for instance, particle-swarm optimization (PSO) and differential evolution (DE)—are used to optimize the allocation of information granularity, and coverage and specificity criteria are used to guide the optimization. A series of experimental studies is reported which offers a comprehensive overview of the underlying realization and performance of the granular fuzzy rule-based models.

Preface

The work conducted in this thesis was performed by Xingchen Hu under the supervision of Professor Witold Pedrycz. Xingchen Hu was supported by the China Scholarship Council under Grant no. 201306110018.

Chapter 3 of this thesis includes the materials published as X. Hu, W. Pedrycz, G. Wu, X. Wang. "Data reconstruction with information granules: An augmented method of fuzzy clustering." *Applied Soft Computing* 55 (2017): 523-532. I was responsible for experiment development, data collection and analysis, and the manuscript composition. W. Pedrycz was the supervisory author and was involved with the concept formation and manuscript composition. G. Wu was involved with part of coding development. X. Wang was involved with manuscript composition.

Chapter 4 of this thesis includes the materials published as X. Hu, W. Pedrycz, X. Wang. "Comparative analysis of logic operators: a perspective of statistical testing and granular computing." *International Journal of Approximate Reasoning* 66 (2015): 73-90. I was responsible for experiment development, data collection and analysis, and the manuscript composition. W. Pedrycz was the supervisory author and was involved with the concept formation and manuscript composition. X. Wang was involved with manuscript composition.

Chapter 5 of this thesis includes the materials published as X. Hu, W. Pedrycz, X. Wang. "Optimal allocation of information granularity in system modeling through the maximization of information specificity: A development of granular input space." *Applied Soft Computing* 42 (2016): 410-422. I was responsible for experiment development, data collection and analysis, and the manuscript composition. W. Pedrycz was the supervisory author and was involved with the concept formation and manuscript composition. X. Wang was involved with manuscript composition.

Chapter 6.1 of this thesis includes the materials published as X. Hu, W. Pedrycz, O. Castillo, P. Melin. "Fuzzy rule-based models with interactive rules and their granular generalization." *Fuzzy Sets and Systems* 307 (2017): 1-28. I was responsible for experiment development, data collection and analysis, and the manuscript composition. W. Pedrycz was the supervisory author and was involved with the concept formation and manuscript composition. O. Castillo and P. Melin were involved with manuscript composition.

Chapter 6.2 of this thesis includes the materials accepted to publish as X. Hu, W. Pedrycz, X. Wang. "Granular Fuzzy Rule-based Models: A Study in a Comprehensive Evaluation of Fuzzy Models." *IEEE Transactions on Fuzzy Systems* (2016). I was responsible for experiment development, data collection and analysis, and the manuscript composition. W. Pedrycz was the supervisory author and was involved with the concept formation and manuscript composition. X. Wang was involved with manuscript composition.

Chapter 6.3 of this thesis includes the materials published as X. Hu, W. Pedrycz, X. Wang. "From fuzzy rule-based models to their granular generalizations." *Knowledge-Based Systems* 124 (2017): 133-143. I was responsible for experiment development, data collection and analysis, and the manuscript composition. W. Pedrycz was the supervisory author and was involved with the concept formation and manuscript composition. X. Wang was involved with manuscript composition.

Chapter 7 of this thesis includes the materials accepted to publish as X. Hu, W. Pedrycz, X. Wang. " Development of Granular Models Through the Design a Granular Output Spaces." *Knowledge-Based Systems* 134 (2017): 159-171. I was responsible for experiment development, data collection and analysis, and the manuscript composition. W. Pedrycz was the supervisory author and was involved with the concept formation and manuscript composition. X. Wang was involved with manuscript composition.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor, Professor Witold Pedrycz. It has been an honor to be his Ph.D. student, for his patience, diligence, motivation, and immense knowledge. I appreciate all his contributions of time, ideas, and guidance to make my Ph.D. experience productive and stimulating. I could not have imagined having a better supervisor for my Ph.D. study.

I would acknowledge National University of Defense Technology, and Professor Hongxu Ma, Professor Xin Xu, who recommended and encouraged me to study abroad. I also acknowledge the supporting grant from the China Scholarship Council under Grant no. 201306110018. So that I can enjoy a wonderful experience in Canada and fully feel the Canadian culture.

I am also grateful to the professors, who have provided me interesting courses and brilliant suggestions during my Ph.D. study, such as Professor Petr Musilek, Professor Marek Reformat, Professor Lukasz Kurgan, Professor Dale Schuurmans, Professor Rich Sutton, etc. I appreciated your time to share their knowledge and appreciation of science. I would also thank the examiners of my PhD final defense for their brilliant comments and suggestions.

I would like to thank my fellow lab mates Dr. Guohua Wu, Dr. Orion Fausto Reyes-Galaviz, Dr. Shiping Wang, Tan Bri, Jinbo Li, Wentao Li for their collaboration and stimulating discussions. Besides, I am fully grateful to my great friends Keyu Wu, Fudong Li, Shengjun Huang, Chao Ren, Jing Yan, YuanYuan, Fanchi Meng, Yining Fang, Ziling Wei, Guoxin Li, Zhongjie Zhang, Xin Cui, Huangchao Yu, Zhixi Yang, and many more. Being with you, I could have the feeling of home in a foreign country. In addition, a very special gratitude goes out to my lovely friends Jun Yan, Jing Tian and Xiaoyu Tan to share memorable experiences and stories across the Pacific Ocean.

Furthermore, I am fully indebted to my parents and my whole family for all their love and encouragement. Thanks for all the support they have provided me over the years, which was the greatest gift in my life. Also, Ling Cao deserves some special thanks for the company in many years regardless the distance and has been like a family member to me.

Last but not least, I would appreciate University of Alberta for its vibrant and supportive learning environment and all the staffs who helped me before for their kindness and cordiality.

Xingchen Hu

University of Alberta

September 2017

Table of Contents

List of Tables

List of Figures

List of Symbols

$\boldsymbol{x}_k$	k-th input data
y_k	k-th output data
$\boldsymbol{v}_i$	i-th prototype in input space
w_i	i-th prototype in output space
U	partition matrix
c	number of clusters
m	fuzzification coefficient
N	the number of data
σ_j	the standard deviation of the j-th variable
A_i	membership function of the i-th fuzzy set in the input space
B_i	membership function of the i-th fuzzy set in the output space
$\boldsymbol{a}$	vector of parameter of fuzzy models based on clustering
$\hat{y}$	output of models
W	weight matrix
X	granular input
Y	granular output
ε	level of information granularity
cov	coverage
sp	specificity
V	global performance of granular fuzzy rule-based models

Chapter 1

Introduction

Fuzzy set theory has witnessed an impressive growth since its inception when Zadeh published the first paper on fuzzy sets [1]. Fuzzy theory is an attractive research direction, as it is the one of the most comprehensible and acceptable of perspectives for describing, understanding and exploring the world by mimicking the habit of human cognition and behavior. Commonly, the feelings and words of human beings expressed in the natural language usually have un-sharp (fuzzy) boundaries. One significant contribution of fuzzy theory is to provide a principal for describing things on a basis of gradation and pluralism rather than via binarization [2]. Fuzzy models (usually encountered as fuzzy rule-based models) are developed originating from fuzzy theory to describe nonlinear systems. They establish a kind of inference systems that imitates the reasoning methods of humans. They have been studied for decades and produced many diverse architectures and design approaches, as for example in [3]-[10]. In the existing diversity of fuzzy models, fuzzy logic theory plays a vital and unique role by forming a human-centric computing framework [11].

As time goes on, fuzzy logic theory has evolved into a more advanced perspective: information granularity, to represent and recognize the abstraction of information as the observation method of humans. Generally, we used to have two broad approaches to reasoning: inductive and deductive methods. In induction, we tend to recognize patterns by grouping them together as information entities according to their internal or external similarity and then developing them into general conclusions or theories. In deductive reasoning, we use inductively obtained results to process information which is usually abstract, fuzzy, concise and imprecise, and we are willing to accept that the reasoning outcomes are vague if it contains useful information. These forms of abstract information entities are termed information granules [12]-[16]. In this perspective, the numeric data can be presented at various levels of resolution or scales, and operated more efficiently. Thus, in recent literature, researchers are interested in realizing and processing data at an abstract or extended level rather than crisp form when they handle issues in the fields of machine learning and data mining [17]-[22].

To use information granules, and to help users understand them, Granular Computing [23][24][26]-[28] and granular modeling [25][31][32][33][36]-[38] have been growing quickly as a paradigms of information processing for information granules in the domain of human-centric systems [28]. In Granular Computing, usually, values, variables, and model systems are granulated in an abstract way, as the manner of human cognition [29]-[37][39]. Among the existing possible perspectives in Granular Computing, we may regard the information granules as interval sets, fuzzy sets, rough sets [40], shadowed sets [41], probabilistic sets, etc. Regardless of how we refer to them, it is a beneficial challenge to set up and design a comprehensive framework to develop and process information granules with an appropriate methodology.

Fuzzy rule-based models are presented as a well-structured framework for processing information granules because of its ability to handle numeric and semantic information in one system. Furthermore, fuzzy rule-based models are able to represent and exploit the knowledge acquired from data learning processes or experts' experiences, and the models can be interpreted by humans as well. Therefore, we construct granular models based on fuzzy rule-based models to develop an advanced way to describe complex nonlinear system. Consequently, granular fuzzy rule-based models underlie the remarkable human ability to engage in rational reasoning when information is imprecise, uncertain, partially known and partial true [23].

1.1 Motivation

One of the most common features of fuzzy models is that they produce numeric results [42], and that the design of the models is carried out at the numeric level guided by a performance index that is typical for numeric models. However, there are no ideal fuzzy models that fully capture (coincide with) all numeric experimental data. In other words, we cannot without any error capture all the inputs that form the output data by the outputs of the model [43]. This is due to the error that comes from almost everywhere in such a model: e.g., the deficiencies of the model's design and the inaccurate parameters of the model. Even the inputs of the model are possibly unreliable (due to measurement error). Thus, error and noise in numeric values are almost inevitable when we develop and use a numeric model. Even more advanced fuzzy models such as those that exploit Type-2 or interval-valued fuzzy sets are susceptible to (encounter) the same error. Moreover, to a certain extent, there is no expectation that minute variations in models could have a significant influence on resulting outputs. Do we indeed care about the difference between the outside temperature is 24 °C and 26 °C? An over-emphasized numeric facet of processing is somewhat counter-intuitive and does not lie in the spirit of any fuzzy processing and system modeling. In general, excessive design effort is wasted, as the model is assessed as a purely numeric construct in the end.

With Granular Computing, we give up obtaining precise numeric models and focus instead on making the models to granular forms, which produce granular outputs capturing target or crucial information of the experimental data. This study presents our efforts toward the granulation of data and fuzzy rule-based models. The fuzzy rule-based models are made granular by admitting and allocating granular parameters at a certain level of granularity so that the granular output of the model can embrace (cover) the target data as much as possible and are as specific as possible [44]-[50]. The resulting granular fuzzy models offer higher tolerance to data noise and modeling errors, and help to produce results that are of practical relevance. The main characteristic of granular fuzzy modeling is the veracious, interpretable, and semantically-oriented transparency of the developed constructs [51][52].

It is obvious that the performance index used to evaluate the granular fuzzy model should not manifest as numeric constructs such as the RMSE or the like because it involves information granularity. Therefore, we turn to two pertinent performance indexes that information granules are guided to form—namely, coverage and specificity criteria—to evaluate and analyze the performance and guide the optimization of the granular fuzzy model.

1.2 Objective and originality

As illustrated in Figure 1.1, the fundamental objective of this study is to establish a comprehensive, systematic method with which to handle the information granules obtained from data based on the fuzzy rule-based model framework. First, we design the granulation and degranulation scheme transforming information granularity from/to numeric data sets, and its augmented approach. Second, we investigate several commonly known logic operators that are used in fuzzy modeling and granular fuzzy modeling. Afterwards, we design a series of new development strategies for granular fuzzy rule-based models by admitting and allocating a certain level of information granularity around numeric values. As a general architecture of fuzzy models can be highlighted three main functional modules, that is: input interface, processing module and output interface, our proposed granular fuzzy rule-based models could be classified into three groups: granular input space of the models, granular processing modules of the models, and granular output space of the models. The performance of granular fuzzy rule-based models is evaluated by two pertinent criteria coverage and specificity. From a methodological point of view, one can stress that granular modeling delivers a successive layer of system modeling. The approach advocated here builds upon an already constructed numeric model and makes it better aligned with the system under consideration by involving the concept of information granules.

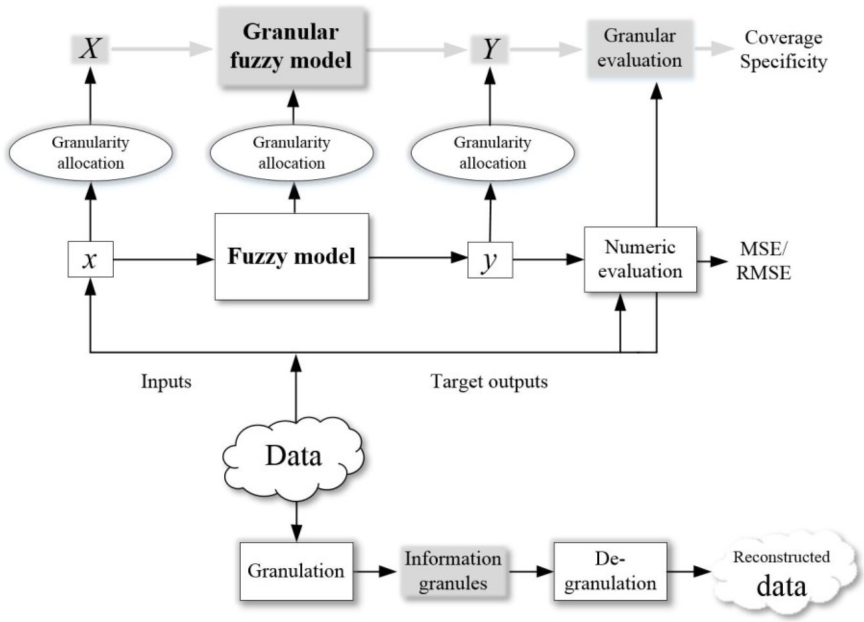


Figure 1.1. A general roadmap of the research

The originality of the research in this thesis is mainly as follows.

(1) We develop a novel augmented mechanism of data reconstruction to reduce the deterioration of a granulation-degranulation scheme, which is of benefit for the transformation of numeric data and information granules.

(2) We design mechanisms of allocation of information granularity across the components of the fuzzy rule-based model in several new strategies which serve as sound vehicles for improving the performance of the model being sought at the higher level of abstraction of information granules.

(3) We provide comprehensive criteria to assess and analyze the performance of the granular fuzzy rule-based models to cope with different strategies.

Overall, the proposed study realizes a new original framework that emphasizes the nature of a granular fuzzy model being regarded as an enhancement of the original fuzzy model in an advanced perspective.

1.3 Organization

The thesis is structured into the following chapters:

To make the study self-contained and easy to follow, Chapter 2 covers some useful background ideas involved in this study, including fuzzy clustering, fuzzy rule-based modeling, statistical analysis, evaluation criteria of information granules, and population-based optimization methods.

Chapter 3 proposes an augmentation mechanism for the generic data reconstruction approach by introducing transformation mapping of the originally produced partition matrix and by setting up an adjustment mechanism modifying a localization of the prototypes to enhance the quality of reconstruction.

Chapter 4 poses a question whether two t-norms produce distinct results. The problem is formally expressed as a certain kind of hypothesis testing in which a null hypothesis concerns the equality of medians of membership grades produced by two triangular norms. In the sequel, we introduce a concept of granular t-norms and discuss an idea of the granular equivalence of logic operators. This study is an indispensable reference for selecting logic operators for fuzzy modeling and granular modeling.

Chapter 5 introduces a concept of a granular input space in fuzzy rule-based modeling and develops an algorithmic framework that supports an optimization of the specificity of the model exposed to granular inputs data. For illustrative purposes, the study is focused on information granules that are formalized in terms of intervals. However, the proposed approach is equally relevant for other formalisms of information granules.

Chapter 6 discusses the concepts and developments of granular fuzzy rule-based models in the processing module of Takagi-Sugeno fuzzy rule-based models. We present augmentation for fuzzy models by forming information granules around the numeric values of the parameters and constructions of the models, and show how different protocols of allocating information granules lead to improve the performance of granular models.

Chapter 7 proposes a method of granular output space and develops an optimization process for the allocation of information granularity across this space. We endow the output space with a mechanism for the optimal allocation of information granularity, which is to say that the numeric results formed by the original model are augmented by interval-information granules whose level of information granularity is determined by the developed mechanism.

Chapter 8 draws the main conclusions of this thesis and lists several promising directions for future research.

Chapter 2

Background

This chapter briefly covers some useful prerequisites that make this study self-contained and easy to follow.

2.1 Fuzzy Clustering

Data clustering is a process of grouping data instances from original data sets into a number of clusters, so that the features depending on the nature of the data are as similar as possible. In granular computing, clustering is one of the approaches most commonly used to generate information granules. For instance, depending upon the nature of the underlying clustering algorithm, the information granules produced arise as sets, fuzzy sets or rough sets. In fuzzy clustering, the distinctions of each datum are measured in the terms of the membership grades. In particular, fuzzy C-means (FCM) [53] along with its numerous extensions [54]-[60], is one variety of the clustering method most frequently used in the formation of information granules.

The generic version of the FCM algorithm minimizes the following objective function *OF* coming in the following form

$$OF = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^{m} \left\| \boldsymbol{z}_k - \boldsymbol{r}_i \right\|^2 \tag{2.1}$$

where $\boldsymbol{z}$ is the data to be clustered containing N data instances, c is a predetermined number of clusters, $u_{ik} \in [0, 1]$ is the elements of partition matrix and $\sum_{i=1}^{c} u_{ik} = 1$, m is a fuzzification coefficient that is usually greater than 1, $\boldsymbol{r}_i$ is prototypes obtained by clustering, and ||.|| is denoted the weighted Euclidean distance expressed as

$$\|\boldsymbol{z}_k - \boldsymbol{r}_i\|^2 = \sum_{j=1}^{n} \frac{(z_{kj} - r_{ij})^2}{\sigma_j^2} \tag{2.2}$$

where σ_j is a standard deviation of the j-th variable of the multivariable data. Obviously, the choice of the distance function impacts the geometry of the clusters and this entails modeling capabilities supported by the ensuing rule-based models.

The objective function *OF* is minimized iteratively by updated the prototypes and the partition matrix successively. In each iteration, the entries of $\boldsymbol{r}_i$ and the k-th data instance is represented in terms of the membership grades in the partition matrix. The prototypes and each element in the partition matrix are calculated as follows, respectively,

$$\boldsymbol{r}_i = \frac{\sum_{k=1}^{N} u_{ik}^m \boldsymbol{z}_k}{\sum_{k=1}^{N} u_{ik}^m} \tag{2.3}$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|\boldsymbol{x}_k - \boldsymbol{r}_i\|}{\|\boldsymbol{x}_k - \boldsymbol{r}_j\|} \right)^{2/(m-1)}} \tag{2.4}$$

The results of clustering of the data partitioned into c clusters come in the form of the partition matrix $U = [u_{ik}]$, $i = 1, 2,\ldots, c$; $k = 1, 2,.., N$ and a collection of prototypes $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, …, $\boldsymbol{r}_c$.

As to the input-output pairs of data $(\boldsymbol{x}_k, y_k)$, $k = 1, 2,\ldots, N$ are concatenated data $\boldsymbol{z}_k = [\boldsymbol{x}_k\ y_k]$. The FCM clustering is completed in this data space. Thus, the prototypes $\boldsymbol{r}_i$ produced by the FCM are expressed as $\boldsymbol{r}_i = [\boldsymbol{v}_i\ w_i]$, where $\boldsymbol{v}_i$ and w_i are the prototypes positioned in the input and output spaces, respectively.

2.2 Fuzzy Rule-based Model Systems

The fuzzy rule-based modeling is an approach to using fuzzy logic to describe and handle complex nonlinear relationships by formulating if-then rules that are overlapped through input and output space and contain extractions of knowledge in the following form:

$$\text{If } \textit{antecedent proposition} \text{ then } \textit{consequent proposition} \tag{2.5}$$

A fuzzy rule-based model system is a collection of rules with a certain model structure. In particular, Takagi-Sugeno (TS) [3] and Mamdani [4] fuzzy rule-based structures have become much renowned over the years. In the Mamdani structure, both the antecedent and consequent are linguistic; in TS models, on the other hand, the antecedent of the model is described by linguistic expressions and the consequents are numeric. This study concentrates on a fuzzy TS, multi-input, single-output (MISO) system. The rationale behind this choice is as follows: a TS fuzzy rule-based model can be regarded as a combination of linguistic and mathematical function modeling, and the output of such a model is crisp and valid. Therefore, TS fuzzy models are commonly encountered in fuzzy modeling and come with a great deal of well-established design practices (quite commonly engaging techniques of evolutionary optimization) and applications, thereby demonstrating their usefulness and relevance. For example, see [61]-[65].

The generic TS fuzzy model yields an aggregation of a collection of fuzzy rules, which are expressed as,

$$\text{The } i\text{-th rule: If } \underbrace{x_1 \text{ is } A_{i,1} \text{ and} \ldots x_n \text{ is } A_{i,n}}_{\text{condition part}} \text{ then } \underbrace{y_i = f_i(x_1, \ldots, x_n)}_{\text{conclusion part}} \tag{2.6}$$

where i = 1, 2,…, c, and c is the number of rules, $\boldsymbol{x}$ is a n-dimensional input variable. A is the membership function (fuzzy sets) respecting to the input variable. The membership functions can be estimated relying on expert experience or by admitting a data driven approach. In the second alternative, clustering algorithm plays an important role to generate clusters (prototypes) and determine the rule structure of the model [66]-[68]. In the conclusion part, y_i is the output of the i-th rule described by some local function $f_i(\boldsymbol{x})$, which is typically regarded as a linear function or simply as some constant values.

When arranging all the rules together involving their condition parts, the output of the model is aggregated by taking the weighted average of the output of each rule.

$$\hat{y} = \frac{\sum_{i=1}^{c} A_i(\boldsymbol{x}) f_i(\boldsymbol{x})}{\sum_{i=1}^{c} A_i(\boldsymbol{x})} \tag{2.7}$$

If we consider implementing FCM clustering to form the membership function (fuzzy set), it holds that $\sum_{i=1}^{c} A_i(\boldsymbol{x}) = 1$. The output of the fuzzy models can be rewritten in the following expression,

$$\hat{y} = \sum_{i=1}^{c} A_i(\boldsymbol{x}) f_i(\boldsymbol{x}) \tag{2.8}$$

In the original TS fuzzy model, $f_i(\boldsymbol{x})$ in the conclusion parts of the rules is adopted as a linear function as follows,

$$f_i(x_1, x_2, ..., x_n) = p_0 + p_1 x_1 + p_2 x_2 + ... + p_n x_n \tag{2.9}$$

In some cases, for the sake of simplification, p_1, p_2, …, p_n can be defined as zeros, so that the function is simplified as a constant value. The parameters vector $\boldsymbol{p}$ in the linear function is identified by using the Least square estimation approach as calculated as follows.

$$\boldsymbol{p} = \left(\boldsymbol{x}^T \boldsymbol{x}\right)^{-1} \boldsymbol{xy} \tag{2.10}$$

Another alternative of fuzzy modeling approach is involving the prototypes obtained from clustering to the local function. Namely, $f_i(\boldsymbol{x})$ stands for a local linear function interpreted as a hyperplane governed by the following expression,

$$f_i(\boldsymbol{x}) = w_i + \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{v}_i) \tag{2.11}$$

where $\boldsymbol{v}_i$ is a cluster (prototype) capturing the location of the rule in the input space $\boldsymbol{R}^n$ and w_i is the corresponding value in the output space, then

$$\hat{\boldsymbol{y}} = \sum_{i=1}^{c} A_i(\boldsymbol{x})\left(w_i + \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{v}_i)\right) \tag{2.12}$$

Let us introduce some auxiliary notation as shown below,

$$\boldsymbol{\Psi} = A_i(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{v}_i) \tag{2.13}$$

$$\boldsymbol{\Theta} = \sum_{i=1}^{c} A_i(\boldsymbol{x}) w_i \tag{2.14}$$

Then the above model is concisely described in the form

$$y = \boldsymbol{\Theta} + \sum_{i=1}^{c} \boldsymbol{a}_i^T \boldsymbol{\Psi} \tag{2.15}$$

We introduce the following concise notation,

$$\widetilde{\boldsymbol{p}} = \left[y_1 - \Theta_1, y_2 - \Theta_2, \ldots, y_N - \Theta_N\right]^T \tag{2.16}$$

In the sequel, the parameters of the model are arranged into the *cn*-dimensional vector

$$\boldsymbol{a} = [a_{11},\ a_{12},\ ...,\ a_{1n},\ a_{21}, a_{22},\ ...,\ a_{2n},\ ...,\ a_{c1},\ a_{c2},\ ...,\ a_{cn}]^T \tag{2.17}$$

Furthermore, the data are structured in the matrix format

$$\tilde{\boldsymbol{\Psi}} = \begin{bmatrix} \Psi_{11} & \Psi_{12} & \cdots & \Psi_{1c} \\ \Psi_{21} & \Psi_{22} & \cdots & \Psi_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_{N1} & \Psi_{N1} & \cdots & \Psi_{Nc} \end{bmatrix} \tag{2.18}$$

Then

$$MSE = \sum_{k=1}^{N} \left(y_k - \Theta_k - \sum_{i=1}^{c} \boldsymbol{a}_i^T \Psi_{ki} \right)^2 = \left(\tilde{\boldsymbol{p}} - \tilde{\boldsymbol{\Psi}} \boldsymbol{a} \right)^T \left(\tilde{\boldsymbol{p}} - \tilde{\boldsymbol{\Psi}} \boldsymbol{a} \right) \tag{2.19}$$

Minimizing (2.19), we estimate the parameters $\boldsymbol{a}$ by the Least square estimation as,

$$\boldsymbol{a}_{opt} = \left(\tilde{\boldsymbol{\Psi}}^T \tilde{\boldsymbol{\Psi}} \right)^{-1} \tilde{\boldsymbol{\Psi}}^T \tilde{\boldsymbol{p}} \tag{2.20}$$

2.3 Statistical Analysis

Statistical comparison is used in these studies to determine if the difference between two methods is statistically significant. We involve two methods here: One is a *t*-test [69]; the other is a non-parametric Mann-Whitney-Wilcoxon test [70]. The *t*-test is a commonly used approach and is relatively easy to understand and perform. However, it is not suitable for all comparisons. For instance, if the data for either test is not normally distributed, then a different test should be employed: e.g., the Mann-Whitney-Wilcoxon test.

The t-test is used to verify that if the means of two sequences of methods is essentially different from one another [71]. We assume the two considered sequences x_1, x_2, …, x_{N1} and y_1, y_2, …, y_{N1} are normally distributed, and the standard deviation of the two sequences Δx and Δy are the same too. The t-values is calculated as following formula,

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{(N_1 - 1)\Delta x^2 + (N_2 - 1)\Delta y^2}} \sqrt{\frac{N_1 N_2 (N_1 + N_2 - 2)}{N_1 + N_2}} \tag{2.21}$$

where $\bar{x}, \bar{y}$ are the mean of the two sequences, respectively.

The null hypothesis H_0 is formulated as follows

$$H_0\text{: the mean of the two sequences are equal} \tag{2.22}$$

For a given confidence level α and $\gamma = N_1 + N_2 - 1$, we compare the t-values with the value $t_{\alpha,\gamma}$ in t-test table. If the following condition satisfies,

$$|t| > t_{\alpha,\gamma} \tag{2.23}$$

We reject the hypothesis and conclude that the means of the two sequences are different, and we can also say that in this case there is a statistically significant difference between the sequences.

The nonparametric Mann–Whitney-Wilcoxon test is more efficient on non-normal distributions than the t-test, and is nearly as efficient as the t-test for normal distributions [72]. The null hypothesis H_0 is formulated as follows

$$H_0\text{: the medians of the two sequences are equal} \tag{2.24}$$

We first determine the mean E_0 and variance var_0 of the sum of the ranks Γ assigned to the first group of samples. They are calculated as follows,

$$E_0 = \frac{N_s(N_1 + N_2 + 1)}{2} \tag{2.25}$$

where $N_s = \min(N_1, N_2)$ and N_1, N_2 are the number of two samples.

$$var_0 = \frac{N_1 N_2 (N_1 + N_2 + 1)}{12} \tag{2.26}$$

$$var_0 = \frac{N_1 N_2}{12}\left[N_1 + N_2 + 1 - \frac{\sum_{j=1}^{g}(tie_j - 1)tie_j(tie_j + 1)}{(N_1 + N_2)(N_1 + N_2 - 1)}\right] \tag{2.27}$$

(2.27) is used for large-sample approximation (the number of instance is larger than 20), where g denotes the number of tied groups and tie_j is the size of tied group j. The term tie is used in connection with rank order, when some values we test are the same, they are put together as a tied group. Tied observation is helpful for making the test results more exact for large-sample. The statistic of interest reads as follows

$$\Gamma^* = \frac{\Gamma - E_0}{\sqrt{var_0}} \tag{2.28}$$

Γ^* has an asymptotic normal distribution, N (0, 1). If $|\Gamma^*| \geq z_{\alpha/2}$, we reject the hypothesis at the α (usually $\alpha = 0.05$) level of significance and the conclusion is: “The two sequences are significantly different”. Otherwise, we do not reject the hypothesis.

2.4 Evaluation Criteria of Information Granules

Commonly, the design of fuzzy models is guided by a numeric-driven performance index (e.g. RMSE or mean-squared error, MSE). This means that, in spite of using fuzzy sets as integral architectural components, fuzzy model outputs are deemed numeric and their performance is evaluated on the basis of this numeric manifestation of the fuzzy model. As mentioned in the introduction, the information granules are constructed in various formalisms: intervals, fuzzy sets, rough sets, etc. For example, the granule generated in a Mamdani fuzzy model is represented in a fuzzy-sets form. The numeric performance of fuzzy models implies that they are legitimately compared with other numeric models, such as neural networks, and that in such cases they may exhibit lower quality. These methodological considerations concerning the evaluation of performance do not seem to have full justification.

To evaluate the various granules, a principle of justifiable information granularity is introduced in [44]. The term "justifiable" pertains to two requirements: (i) highly accumulated numeric experimental evidence contains in the granules, and (ii) specific enough to define an articulated semantics (meaning) of the granules. Guided by the principle of justifiable, two criteria are frequently encountered in literatures to assess the performance of the information granularity [45]-[48], namely coverage and specificity.

2.4.1 Coverage criterion

A fundamental criterion used to assess the performance of the granular model concerns coverage. In essence, coverage expresses an extent to which information granule produced by the granular model Y_k "covers" target output y_k, viz. the experimental datum is represented by the result produced by the model. Considering a collection of data, the overall coverage is expressed as the following sum

$$cov = \frac{1}{N}\sum_{k=1}^{N}\text{incl}\left(y_k, Y_k\right) \tag{2.29}$$

Evidently, the higher the coverage, the better the model with respect of its modeling capabilities. The inclusion predicate (incl) has to be specified depending upon the formal way in which information granule Y_k has been formalized.

If Y_k comes in the form of a certain interval, the inclusion predicate is expressed as follows,

$$\text{incl}(y_k, Y_k) = \begin{cases} 1 & y_k \in Y_k \\ 0 & \text{otherwise} \end{cases} \tag{2.30}$$

In a nutshell, by using the above performance measure one counts the number of instances of inclusion of data y_k in the granular output of the model and returns average value computed over all data. In an ideal situation, cove returns 1, viz. all data are "covered" by the granular output.

In case of Y_k being fuzzy sets, the inclusion operation returns a membership value of y_k in Y_k, namely $Y_k(y_k)$. Assume the fuzzy sets are described by fuzzy numbers $[l_k^*, y_k^*, u_k^*]$, and the membership function is expressed by left- and right- hand bounded function f_k and g_k as shown in Figure 2.1.

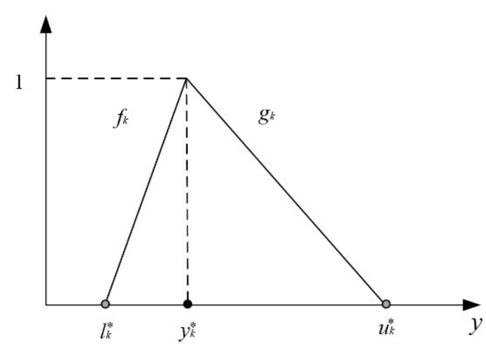


Figure 2.1. Examples of membership functions.

In virtue of the existing membership function Y_k, the coverage is computed as the membership grade of the outcome fuzzy sets.

$$\text{incl}(y_k, Y_k) = \begin{cases} f_k(y_k) & \text{if } l_k^* \le y_k < y_k^* \\ g_k(y_k) & \text{if } y_k^* \le y_k \le u_k^* \\ 0 & \text{otherwise} \end{cases} \tag{2.31}$$

2.4.2 Specificity criterion

The coverage plays an important role, however one has to take into consideration the quality of the granular output. It is expressed in the form of information specificity. This measure evaluates how specific (detailed) a certain information granule Y_k is. In general, by specificity of Y_k , $sp(Y_k)$ we view a functional defined over Y_k satisfying the condition of monotonicity: if $Y_k \subset Y_k'$ then $sp(Y_k) \geqslant sp(Y_k')$, and the boundary condition $sp(\{y\}) = 1$.

When considering an interval form of Y_k, the shorter the interval, the higher its specificity becomes. In a limit case, once Y_k reduces to a single point, the specificity attains its maximal value of 1. For example, one among possible alternatives using which the specificity can be expressed comes in the following form

$$sp = \frac{1}{N}\sum_{k=1}^{N} \exp\left(-\left|y_k^+ - y_k^-\right|\right) \tag{2.32}$$

Obviously, instead of the exponential function used above, one could consider any continuous decreasing function of the length of the interval.

When Y_k is encountered as fuzzy sets, the calculations specificity of Y_k involves evaluating the size of the fuzzy sets, namely the area under the membership function curves. Intuitively it relates to the "size" of the fuzzy set. If the fuzzy set is a single-element entity, its specificity attains 1. The larger the size, the lower the specificity. The calculations specificity of the fuzzy set Y_k involves the left- and right- hand side parts of the membership function ($f_k(\boldsymbol{x})$ and $g_k(\boldsymbol{x})$) with applying a series of its α-cuts to evaluate the size of the output fuzzy sets,

$$sp_k = 1 - \int_0^1 \frac{\left(y_k - f_k^{-1}(\alpha)\right) + \left(g_k^{-1}(\alpha) - y_k\right)}{range} d\alpha = 1 - \int_0^1 \frac{g_k^{-1}(\alpha) - f_k^{-1}(\alpha)}{range} d\alpha \tag{2.33}$$

where *range* is computed the range of target outputs, as $y_{max} - y_{min}$. In fact, the part of expression in the integral formula $\int_0^1 g_k^{-1}(\alpha) - f_k^{-1}(\alpha) d\alpha$ is calculated the area under the left- and right- hand side membership functions.

The performance of the information granularity and granular model in this research is assessed by considering the criteria of coverage and specificity.

It is worth stressing that these two values depend upon the predetermined level of information granularity ε. To form a measure being independent from this level and produce a global characterization, we form evaluation indicator for various values of ε. This fact could be stressed by using the alternative notation $cov(\varepsilon)$ and $sp(\varepsilon)$.

2.4.3 The overall performance indicator

The performance of the granular model is assessed by considering the criteria of coverage and specificity. We strive to simultaneously maximize the coverage and the specificity. Apparently, these two measures are in conflict. Higher coverage values imply lower specificity values.

For different situations and concerns, the indicator can be considered as coverage, specificity, or both coverage and specificity criterions. In coverage coordinate, specificity coordinates or coverage-specificity coordinates, we form indicator for various values of ε, respectively,

$$Q = cov(\varepsilon) \tag{2.34}$$

$$Q = sp(\varepsilon) \tag{2.35}$$

$$Q = cov(\varepsilon)sp(\varepsilon) \tag{2.36}$$

For some given ε, the optimization of (2.34)-(2.36) results in an optimal allocation of information granularity.

When the objective is to maximize (2.36) through an allocation of information granularity ε (design asset), viz. by designing (optimizing) the transformation module, coverage, specificity and Q are related to the allocation of information granularity ε, which is illustrated as Figure 2.2. The plots display the general characteristics of the components of the objective function. It is also noticeable that the coverage and specificity are in conflict and have to be compromised.

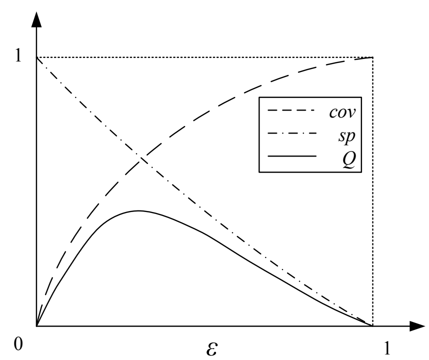


Figure 2.2. Example relationship between coverage, specificity, V and the allocation of information granularity ε.

In some cases, (2.36) is considered as following form,

$$Q = cov(\varepsilon)\left(sp(\varepsilon)\right)^{\beta} \tag{2.37}$$

where β assuming non-negative numbers is an additional coefficient (weight) that controls the impact of specificity criterion on the objective values. Higher value of β underlines the more significance of specificity. For instance, if $\beta = 0$, the fitness function is only focused on coverage criterion. If $0 < \beta < 1$, the coverage has more effect on the fitness values than specificity. If $\beta = 1$, the coverage and specificity have the same importance. If $\beta > 1$, the fitness function has more influenced by specificity.

The optimization criterion (2.36) can also be modified by focusing on one criterion and requesting that another one satisfies some constraint. For instance, we may optimize coverage and the same time requesting that the specificity does not go beyond some threshold γ, which leads to the problem in the form

$$\text{Maximize}_{\text{allocation of information granularity}}\ cov(\varepsilon) \text{ subject to } sp(\varepsilon) > \gamma \qquad (2.38)$$

where the maximization is expressed in a general fashion (the details will be discussed later in one of the subsequent sections).

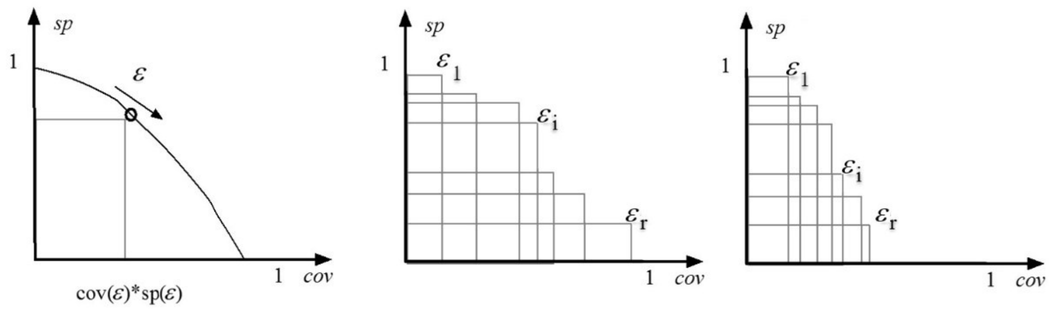


Figure 2.3 Performance of the granular model expressed in the coverage –specificity coordinates: (a) monotonically decreasing values of specificity with the increase of coverage, (b) significant drop in the specificity with some limited increase in coverage at $\varepsilon = \varepsilon_i$, (c) granular model characterized by low AUC value

To develop a global measure of performance being independent from this level and produce a global characterization of the model, the characteristics of the obtained granular model can be displayed in the coverage and specificity coordinates. In the coverage-specificity coordinates we form evaluation for the model for various values of ε and subsequently estimation of an area under the curve (AUC). They deliver a comprehensive insight into the performance of the model and their dependence upon the changes in the values of ε. Several plots, see Figure 2.3, are displayed showing various ways in which increasing values of ε impact the coverage and specificity. For $\varepsilon = 0$, the specificity is 1 while the coverage is practically equal to 0. With the increase of ε, the coverage increases but we pay a price of specificity reduction as the values of this measure are reduced. There could be segments of the curve where coverage still increases not impacting the specificity in a significant measure.

Higher values of AUC indicate that for some values of ε, the granular model "covers" more data yet producing results of higher specificity implying a high quality of the model

$$V = \frac{1}{M}\sum_{\varepsilon} Q(\varepsilon) \tag{2.39}$$

where M is the number of values assumed by ε. In general, if an infinite number of the levels of information granularity is sought, the above expression is replaced by an integral over ε, namely

$$V = \int_0^1 Q(\varepsilon) d\varepsilon \tag{2.40}$$

Refer to Figure 2.3 in which the performance of the granular model (expressed in the *AUC* value) in Figure 2.3 (c) is far lower than the one in Figure 2.3 (b).

Although the result – a granular fuzzy model, may, on surface, exhibit some close resemblance with type-2 fuzzy models present quite commonly in the literature, there are two important differences. First, the design promoted here exhibits two well-delineated phases whereas type-2 fuzzy models are built in a single-step process, which inevitably engages a huge search space (and what implies a huge computing overhead and eventual inefficiency). Second, what is even more important, the evaluation of such models is carried out in a "traditional" manner and this entails the use of the mechanisms of order reduction and a conversion (decoding) of the result into a numeric outcome so that the standard *RMSE* (or any other number-oriented performance index) can be used. In other words, while the enhanced flexibility has been brought to the picture by type-2 fuzzy sets, their potential in system modeling has not been taken advantage of. Type-2 fuzzy model is constructed by being built by the numerically navigated optimization criterion (which involves a numeric manifestation of the model). In contrast, the granular fuzzy model is constructed through the guidance offered by the two measures of performance discussed *en block*, namely coverage and specificity.

2.5 Optimization Methods

The optimization of the granular data and fuzzy model is embarked on the allocation of information granularities regarding to the data or the architecture of the fuzzy model. Generally, the two performance indexes, coverage and specificity, are conflicting, besides, in most cases, the fitness function is nonconvex or not deferential guaranteed. In light of this, some population-based optimization algorithms are resorted, such as the particle swarm optimization (PSO) [73], Differential Evolution (DE) [74] to allocate the information granularities.

2.5.1 PSO algorithm

PSO algorithm is a well-known swarm intelligence algorithm commonly used because of its simplicity and effectiveness for solving complex problem. Each individual in PSO searches the solution space on a basis of its own experience and a collective experience collected so far by the entire swarm. The formulas governing the PSO search concern the position and velocity of the particle, namely $\boldsymbol{pos}_i$ and $\boldsymbol{vel}_i$. The velocity and position of the *i*th particle are updated at each generation during the optimization process,

$$\boldsymbol{pos}_i = \boldsymbol{pos}_i + \boldsymbol{vel}_i \tag{2.41}$$

$$\boldsymbol{vel}_i = \lambda \cdot \boldsymbol{vel}_i + \xi_1 \cdot \boldsymbol{rand}_1 \bullet (\boldsymbol{q}_{best} - \boldsymbol{pos}_i) + \xi_2 \cdot \boldsymbol{rand}_2 \bullet (\boldsymbol{g}_{best} - \boldsymbol{pos}_i) \tag{2.42}$$

where i stands for in index of the individual particle, and λ is an inertia weight. $\boldsymbol{q}_{best}$ is the personal best solution found by the *i*th particle so far, and $\boldsymbol{g}_{best}$ is the global best solution obtained so far by the entire swarm. ξ_1 and ξ_2 are two acceleration coefficients while $\boldsymbol{rand}_1$

and $\boldsymbol{rand}_2$ are vectors of random numbers coming from the uniform distribution over [0, 1]. The symbol • indicates that the vectors are multiplied coordinate-wise.

2.5.2 DE algorithm

DE algorithm is another popular population-based evolutionary optimization method. The population evolves towards to an optimum solution through a series of evolution operations, such as mutation, crossover, and selection. At the beginning, the population ***pop*** is initialized randomly. Then, in each generation, a mutant vector ***mv*** is generated as follows:

$$\boldsymbol{mv}_i = \boldsymbol{pop}_{q1} + F_i \cdot \left(\boldsymbol{pop}_{q2} - \boldsymbol{pop}_{q3}\right) \tag{2.43}$$

where $q1$, $q2$, and $q3$ $\in \{1, 2, \ldots, NP\}$ are random exclusive integer indexes, NP is the population size and F is the scaling factors positioned in the (0, 2] interval and treated as control parameters. The crossover operation is used to produce a trial vector (denote by ***tv***) by mixing ***mv*** and ***po***,

$$tv_{j,i} = \begin{cases} mv_{j,i} & \text{if rand}[0,1) \le CR \text{ or } j = j_{rand} \\ po_{j,i} & \text{otherwise} \end{cases} \tag{2.44}$$

where rand[0, 1) stands for a random number coming from the uniform distribution within the range [0, 1) and CR ($0<CR<1$) is crossover rate, which controls how many components of ***tv*** are inherited from ***mv***. j_{rand} is an integer index that is selected randomly from the uniform distribution spread over the range [1, D], which guarantees that at least one component of ***tv*** is inherited from ***mv***.

For each variable j in the i-th individual, if the trial vector $tv_{j,i}$ is located beyond the boundaries of the search space [$po_{\min}$, $po_{\max}$], the correction operation is triggered as shown below,

$$tv_{j,i} = \begin{cases} \left(po_{\min,j} + po_{j,i}\right)/2 & \text{if } tv_{j,i} < po_{\min,j} \\ \left(po_{\max,j} + po_{j,i}\right)/2 & \text{if } tv_{j,i} > po_{\max,j} \end{cases} \tag{2.45}$$

Once all the trial vectors have been modified, a selection process determines the survivors for the next generation,

$$\boldsymbol{po}_i = \begin{cases} \boldsymbol{tv}_i & \text{if } Q(\boldsymbol{tv}_i) \le Q(\boldsymbol{po}_i) \\ \boldsymbol{po}_i & \text{otherwise} \end{cases} \tag{2.46}$$

where Q is the fitness function of the optimization.

2.6 Summary

This chapter covers the main approaches and algorithms that are essential to the design of the granular, fuzzy rule-based models in this thesis. Overall, the fuzzy clustering algorithm (in particular, FCM) is used to form numeric, fuzzy rule-based models. The statistical-analysis methods are used to compare the performance of numeric results, and coverage and specificity criteria are introduced to evaluate and optimize the granular fuzzy rule-based models. Some population-based optimization methods (PSO and DE) are implemented to search for the optimal allocation of information granularity.

Chapter 3

Granulation and Degranulation[a]

Information granules are examples of abstract entities delivering a concise and efficient characterization of numeric data at a higher level of abstraction [75][76]. Fuzzy clustering provides a way to describe an imprecise allocation of data to the clusters (by making them granular) and captured by membership functions and generate information granules. In particular, as a fuzzy clustering method based on objective function, FCM becomes a visible technique predominantly because of its simplicity and efficiency [77]. Thus, FCM along with its numerous extensions are a clustering method frequently used in the formation of information granules [66][78]-[80]. In general, clustering realizes information granulation, viz. encoding scheme [81]-[83] by representing any numeric datum in terms of the already constructed information granules. In the FCM algorithm, the structure in the data set is expressed in terms of prototypes (clusters) and partition matrices. Subsequently, data are encoded to information granules with the aid of constructed prototypes and partitions.

The fuzzy clustering techniques are basically focused on the abstraction of the original data. In other words, numeric data is represented by information granules (linguistic data) and described by prototypes and partition matrices. We are also interested in the reconstruction of numeric results on a basis of already constructed information granules. This could be considered as an inverse problem of generic clustering or granulation. In the sequel, the reconstruction of information granules, usually referring to as degranulation or decoding process, returns a numeric result. The concept of granulation-degranulation plays a visible role in Granular Computing, just as fuzzification-defuzzification in fuzzy control systems, and analog-to-digital (A/D) as well as digital-to-analog (D/A) conversion in

[a] A version of this chapter has been published as [128].

digital signal processing [81][83]. In the granulation-degranulation scheme, this leads to some unavoidable deterioration of original data. Practically, the degranulation (reconstruction) is frequently noted in the literature when assessing the performance of clustering by assessing the degranulation deterioration criterion, such as [82][84][85]. Simply put, the lower the degranulation error is, the better performance of the clustering is. In these studies, the reconstruction error depends on the granulation scheme and its parameters. In the context of the FCM method, these parameters involve the number of clusters (information granules) and the value of the fuzzification coefficient. In [81] an impact of these two parameters of granulation procedure on the reconstruction error has been thoroughly investigated and quantified. Thus, the degranulation error is used to determine the optimal values of the number of clusters and fuzzification coefficients for fuzzy clustering. Being cognizant of the centrality of the granulation-degranulation schemes, it is of interest to study further enhancements of the existing schemes by augmenting the degranulation procedure with intent to minimize the associated reconstruction (degranulation) error.

The main objective of this study is to develop an augmented mechanism of data reconstruction to enhance the performance of granulation-degranulation scheme, namely to reduce the deterioration of the reconstruction results. Here, unlike the other methods only aimed at finding the optimal clustering parameters (viz. the number of clusters and the fuzzification coefficient), there are two more ways proposed to reduce the reconstruction losses. First, we introduce a linear transformation of the originally FCM-produced partition matrix to carefully capture high-level associations among the data being clustered. Second, we develop an adjustment mechanism of the prototypes. The merit of the augmented clustering mechanism is to discover more suitable information granules, since both the partition matrix and prototypes are modified. An overall development environment is implemented with the use of population-based optimization.

3.1 Granulation and Degranulation Scheme

Granulation

With the FCM algorithm, the data set is structured and expressed in terms of prototypes and partition matrices. Subsequently, data are granulated or encoded to information granules with the aid of the prototypes and partitions. The granulation of data $\boldsymbol{z}_k$ results in their representation in terms of the membership grades contained in the k-th column of the partition matrix. Each column is a result of the granulation of the corresponding instance of the data, which means a numeric datum $\boldsymbol{z}_k$ is represented in terms of an information granule $\boldsymbol{u}_k$, as described in (2.4).

Degranulation

In the degranulation process, we transform the internal representation of $\boldsymbol{z}_k$ described in terms of information granules into a numeric counterpart (we say that it has been degranulated). The degranulation is realized on a basis of the prototypes and the k-th column of the partition matrix U [81],

$$\boldsymbol{z}_k = \frac{\sum_{i=1}^{c} u_{ik}^{m} \boldsymbol{r}_i}{\sum_{i=1}^{c} u_{ik}^{m}} \tag{3.1}$$

In this expression, each prototype is weighted by the activation level (membership degree) of the corresponding information granules. In virtue of the underlying processing described above the degranulation mechanism relies on the membership grades and the prototypes, see Figure 3.1.

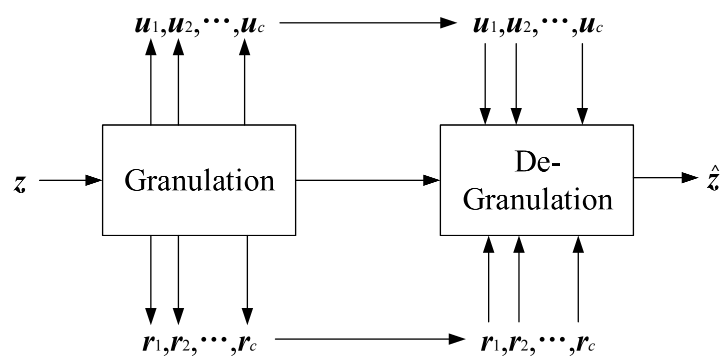


Figure 3.1. General granulation-degranulation scheme

The reconstruction error is quantified in terms of the following performance index

$$PI = \frac{1}{N}\sum_{k=1}^{N}\left\| z_k - \hat{z}_k \right\|^2 \tag{3.2}$$

with the distance ||.|| computed in the same way as in (2.2).

The values of *PI* are impacted by the number of clusters *c* and the fuzzification coefficient *m*. In general, the reconstruction error is a decreasing function of the number of clusters. This is intuitively convincing. The dependence of *PI* upon the values of *m* is not obvious and, in general, becomes data dependent. It has been shown that usually *PI* regarded as a function of *m* shows a certain minimum and quite often the optimal value of *m* is different from the value of 2, which is commonly encountered in the literature.

3.2 An Augmented Granulation and Degranulation Scheme

It is evident that the forming and quantizing of the information granularity within the granulation-degranulation scheme leads to some deterioration of the original data. In the context of granulation and degranulation scheme, the error *V* depends on two parameters: a number of clusters (information granules) and the fuzzification coefficient, which has been thoroughly investigated and quantified in. Here, we develop an augmented mechanism of reconstruction to enhance the performance of granulation-degranulation scheme in two ways by introducing an interactive (interaction) linkage matrix and a mechanism of modification of the FCM-based prototypes as illustrated in Figure 3.2.

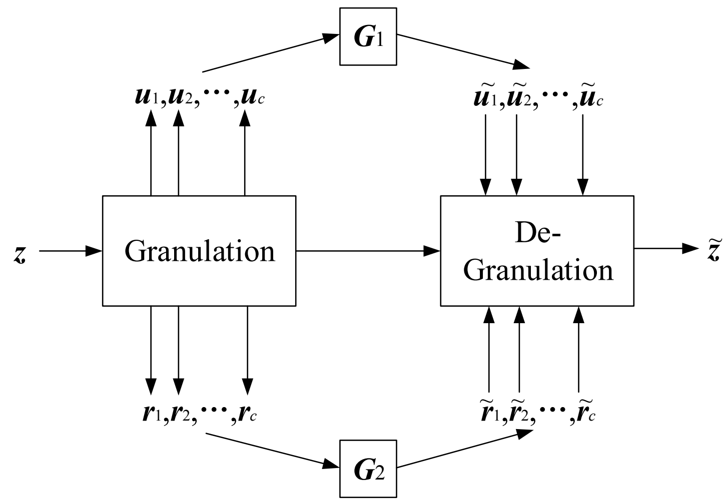


Figure 3.2. Granulation and degranulation: an augmented structure

The objective of the interaction matrix is to modify the partition matrix and allow for an interaction at the level of information granules so that the reconstruction error can be minimized. Prototypes produced by the FCM are in some sense (weighted) averages of the groups of data clustered and as such are obviously not capable of representing the data that are distantly positioned from these prototypes. Possible modifications of the prototypes are considered to increase their dispersion in the data space.

In general, such interaction can be captured in the form of a certain mapping that transforms original information granules, say $\boldsymbol{u}_i$, to new information granules incorporating the linkages (associations) among $\boldsymbol{u}_i$s. Those linkages are realized in the form of a c by c dimensional interaction matrix W_{ij}. Thus, the transformation of the partition matrix is carried out in the form

$$\tilde{u}_{ik} = \sum_{j=1}^{c} W_{ij} u_{jk} \quad i = 1, 2, ..., c \tag{3.3}$$

A result comes as a matrix $\widetilde{U}$, which consists of fuzzy sets $\widetilde{\boldsymbol{u}}_i$ whose membership functions are located at the successive rows of the partition matrix. In other words, $\widetilde{\boldsymbol{u}}_i$ is expressed as $\widetilde{\boldsymbol{u}}_i = G_1(\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_c)$. The interaction matrix realizes the relationship among the elements of the partition matrix in both inhibitory and excitatory manner. The value of W_{ij} is contained within the range [-1, 1].

The prototypes are adjusted in terms of their location in the feature (data) space by admitting an expansion of its individual coordinates. The modification is expressed as $\widetilde{\boldsymbol{r}}_i = G_2(\boldsymbol{r}_i)$, where G_2 stands for the expansion operator of the individual coordinates of the data space. The i-th prototype $\boldsymbol{r}_i$ now becomes $\widetilde{\boldsymbol{r}}_i$ with the coordinates located in the range $[min_j(1-\mu), max_j(1+\mu)]$ where μ is a positive expansion coefficient and $range_j$ stands for the difference between the largest and the smallest value of the j-th variable of the data $x_{1j}, x_{2j}, \ldots, x_{Nj}$, namely $min_j = \min_{k=1,2,\ldots,N} x_{kj}$, $max_j = \max_{k=1,2,\ldots,N} x_{kj}$. The expansion coefficient controls the expansion of the space.

Because of the use of these two transformation mechanisms, the reconstruction mechanism results in the following expression,

$$\widetilde{\boldsymbol{z}}_k = \frac{\sum_{i=1}^{c} \widetilde{u}_{ik}^{m} \widetilde{\boldsymbol{r}}_i}{\sum_{i=1}^{c} \widetilde{u}_{ik}^{m}} \tag{3.4}$$

The objective function is similar as in formula (3.2). It is easy to demonstrate that the objective function is nonconvex and complex with regard to the optimized parameters.

3.3 Experimental Studies

We present a series of experiments involving both synthetic data and several publicly available data, and report on comparative studies by contrasting the performance and the main results produced by the proposed modified reconstruction approach with the results produced by the original method. In the experiments, the data set is split into its training (70%) and testing part (30%). The value of the expansion coefficient μ was set to 0.2.

For a given data set, we select a range of the number of clusters. The granulation scheme was carried out by the FCM algorithm by sweeping across the values of the fuzzification coefficient positioned within the range [1.1, 3] with the step size of 0.1. When running the clustering algorithm, the stopping criterion expressed by the minimum changes of the objective function F is set as 10^{-5}. Then the data are reconstructed as described in Section 2. We record the minimal values of PI associated with the optimal value of the fuzzification coefficient.

We implement DE algorithms, and to compare the performances coming from different optimization algorithm, here we involve two variant DE algorithms, namely, success-history based adaptive differential evolution (SHADE) and SHADE with a linear population size reduction strategy (L-SHADE).

SHADE algorithm is an efficient DE variant, which improves JADE [86] by using a history based parameter adaptation scheme [87]. The mutation strategy used by JADE is current-to-*p*best/1 strategy:

$$\boldsymbol{mv}_i = \boldsymbol{po}_i + F_i \cdot \left(\boldsymbol{po}_{\text{best}}^{p} - \boldsymbol{po}_i\right) + F_i \cdot \left(\boldsymbol{po}_{q1} - \boldsymbol{po}_{\tilde{q}2}\right) \tag{3.5}$$

where $\boldsymbol{po}_{\text{best}}^{p}$ is randomly chosen as one of the top 100*p*% individuals in the current population. $\boldsymbol{po}_{\tilde{q}2}$ is randomly chosen from the current population and the archive. Unlike JADE which generates new control parameter settings based on some distribution around a single pair of parameters μ_{CR} and μ_F, *SHADE* uses a historical memory M_F and M_{CR} to preserve a collection of *CR* and *F* values which have helped *SHADE* to produce promising solutions in the previous generations, and then generate new *CR, F* pairs by directly sampling the parameter space close to one of the preserved pairs. Essentially, the parameters of *SHADE* are adapted by learning previous search experience on-line

$$CR_i = \mathrm{randn}_i\left(M_{CR,rand_i}, 0.1\right) \tag{3.6}$$

$$F_i = \mathrm{randc}_i\left(M_{F,rand_i}, 0.1\right) \tag{3.7}$$

where randn and randc denote normal distribution and Cauchy distribution, respectively. $rand_i$ is an index randomly selected from the uniform distribution over [1, H] whereas H is the memory size. Once the control parameters F_i and CR_i have been assigned for each individual $\boldsymbol{po}_i$ of the population, the corresponding mutant vector $\boldsymbol{mv}_i$ is generated as in (17). In SHADE, each individual $\boldsymbol{po}_i$ has an associated p_i, which is uniformly random generated in [2/NP, 0.2]. The indices $q1$ and $q2$ $\in \{1, 2, \ldots, NP\}$ are randomly selected integer numbers such that they are different from i. After generating mutant vector $\boldsymbol{mv}_i$, the correction and selection operation are similar as in the canonical DE algorithm.

L-SHADE algorithm augments SHADE with a linear population size reduction (LPSR) technique, which enhances the performance of SHADE by continuously reducing the population of the individuals [88]. L-SHADE is the winner of CEC 2014 competition when dealing with real parameter single objective numerical optimization.

The mutation strategy is also *current-to-pbest* as in equation (41), but in L-SHADE, the parameter p is static and its value is set manually. The crossover, correction and selection operations are implemented through (2.43)-(2.46), respectively. As in SHADE, the control parameters F_i and CR_i that succeed to generate better solutions are recorded in historical-memory at each generation, so that they can be adjusted automatically.

Furthermore, a linear population size reduction technique is applied to reduce the size of population in successive generations. The population size in the consecutive generation G+1 is computed as

$$NP_{G+1} = \text{round}\left[\left(\frac{NP_{\min} - NP_{\text{init}}}{NFE_{\max}}\right) \cdot NFE + NP_{\text{init}}\right] \quad (3.8)$$

where $NP_{\min}$ is the possible smallest value such that the evolutionary operators can be implemented, NP_{init} is the initial population size, NFE is the number of evaluations of the fitness at current generation, and $NFE_{\max}$ is the allowed maximal number of fitness evaluations. When $NP_{G+1} < NP_G$, the (NP_G - NP_{G+1}) worst-ranking individuals are removed from the population.

In the sequel, we process the modified reconstruction scheme as described in Section 3. The interaction matrices and the prototypes are optimized by running the DE, SHADE, and L-SHADE algorithms. To fully compare the algorithms, the numbers of used fitness evaluations was set to $D \times 10{,}000$. The parameters settings of the algorithms are completed on a basis of the recommendations available in the literature and by running some trial-and-error experiments. According to the experience gained here, the sound setting of the parameters is listed below:

- DE: $CR = 0.9$, $F = 0.5$ [74], $NP = 100$.
- SHADE: $p = \text{rand}(2/NP, 0.2)$, $NP = 100$, $H = 100$ [87].
- L-SHADE: $p = 0.11$, $H = 6$, the minimal population size is 4, external archive size rate r^{arc} is 2.6 [88], initial $NP = 18D$, the external archive size is set to $NP \times r^{arc}$.

3.3.1 A Synthetic Data Set

The first experiment involves a group (500 instances) of two-dimensional synthetic data, which are generated randomly using normal distributions centered around five prototypes $\boldsymbol{r}_1 = [-7, -6]^T$, $\boldsymbol{r}_2 = [-6, -8.5]^T$, $\boldsymbol{r}_3 = [-0.5, 1]^T$, $\boldsymbol{r}_4 = [8, 4]^T$, $\boldsymbol{r}_5 = [6, -7.5]^T$, The standard deviation of the data in each cluster is set to 4.

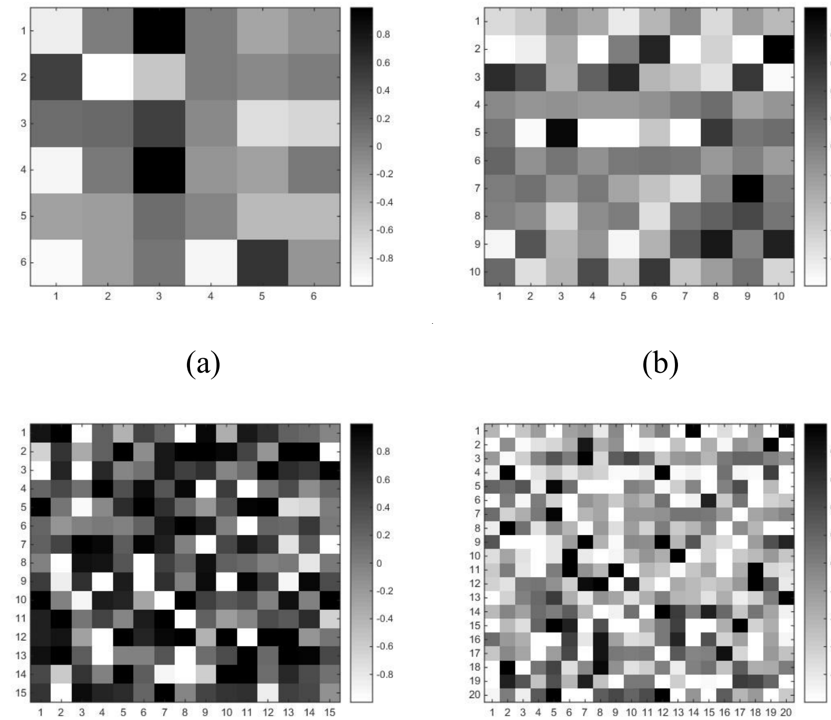
Table 3.1 contrasts the performance produced by the original and improved reconstruction method and then offers a comparative analysis of the optimization performance. In this experiment, the optimal value of the fuzzification coefficient of each model (*c* is selected as 6, 10, 15, 20) is 1.6, 1.7, 1.5 and 1.6, respectively. In general, comparing the results produced by the original and the modified reconstruction approach, we conclude that the improvement provided by the proposed modified method is visible for the most cases both for the training and testing data. Comparing DE, SHADE and L-SHADE algorithms, SHADE and L-SHADE provide better results than DE, and L-SHADE has better performance on testing data. In addition, as expected, the higher the number of clusters, the lower the reconstruction error becomes.

Table 3.1. *PI* for synthetic data sets for selected number of clusters and the corresponding optimal values of the fuzzification coefficient

	Training data				Testing data			
	$c = 6$	$c = 10$	$c = 15$	$c = 20$	$c = 6$	$c = 10$	$c = 15$	$c = 20$
Original	0.2910	0.1607	0.1006	0.0735	0.3128	0.2023	0.1261	0.0943
DE	0.1814	0.0797	0.0559	0.0690	0.2142	0.1148	0.0852	0.0878
SHADE	0.1573	0.0493	0.0401	0.0108	0.1938	0.0817	0.1041	2.8041
L-SHADE	0.1568	0.0556	0.0422	0.0169	0.1656	0.1064	0.0894	0.0459

To visualize the modifications and contrast the performance associated with the proposed methods, we consider the experimental results of L-SHADE as an example and plot the values of the interaction matrix and visualize the changes in the position of the prototypes, refer to Figure 3.3 and Figure 3.4. The interaction matrix and the prototypes have been affected more visibly.

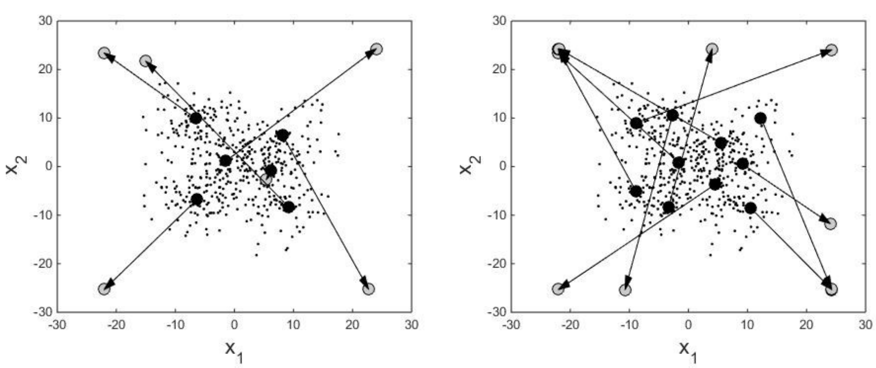
The elements on the diagonal of the matrix are no longer kept close to one, and all the elements show some departure towards negative or positive values. Similarly, most prototypes have been moved visibly from their original locations. Interestingly, when the number of cluster becomes larger, say, $c = 15$ and $c = 20$, some of the modified prototypes being optimized overlap or are positioned close to each other. We note that this could offer some potential to reduce the number of the prototypes, which entails a useful compression effect.



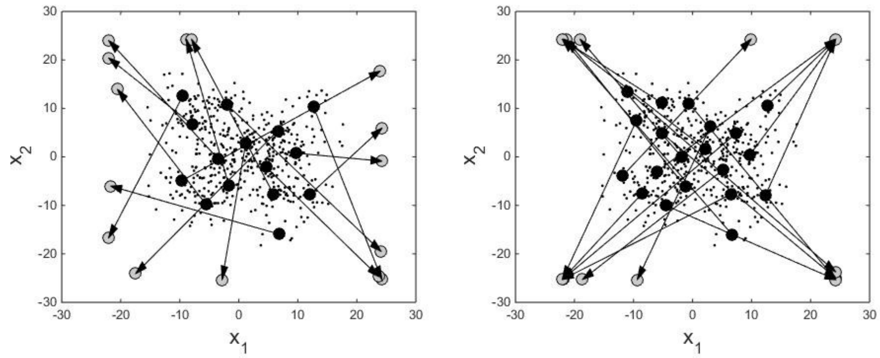
(a) (b)

(c) (d)

Figure 3.3. Plots of interaction matrices optimized by the L-SHADE method: (a) c = 6, (b) c = 10, (c) c = 15, (d) c = 20.



(a) (b)



(c) (d)

Figure 3.4. Plots of the modified prototypes optimized by the L-SHADE method – black circles: original prototypes, gray circles: modified prototypes: (a) $c = 6$, (b) $c = 10$, (c) $c = 15$, (d) $c = 20$.

3.3.2 Publicly Available Data Sets

To demonstrate the usefulness and quantify the performance of the introduced approach, we report on a series of real-world data sets. All the data come from UCI machine learning repository (http://archive.ics.uci.edu/ml/). To make the experiments statistically sound, the 10-fold cross validation was used. The obtained results (the average values of Q and their standard deviations) are contained in Table 3.2. The obtained optimal values of the fuzzification coefficient are listed in Table 3.3. We also compare the four approaches using statistical testing with the results presented in Table 3.4. The two-tailed *t*-test is engaged to test whether the two group of results are statistically different. The significance level of the null hypothesis is set as 0.05. The '+' sign indicates that the differences between the results are significantly different and better. The '-' sign denotes statistically significant difference and worse, whereas 'o' indicates that there is no statistical difference.

Table 3.2. *PI* for real-world data sets for selected number of clusters and the corresponding optimal values of the fuzzification coefficient

	Methods	Training data				Testing data			
		$c = 6$	$c = 10$	$c = 15$	$c = 20$	$c = 6$	$c = 10$	$c = 15$	$c = 20$
Auto-MPG	Original	1.700±0.045	1.205±0.035	0.903±0.027	0.743±0.035	1.707±0.360	1.303±0.360	1.095±0.364	0.933±0.229
	DE	1.515±0.064	1.184±0.257	0.875±0.043	0.720±0.031	1.623±0.235	1.282±0.333	1.058±0.297	0.890±0.211
	SHADE	1.451±0.051	0.914±0.102	0.559±0.083	0.399±0.024	1.522±0.354	1.157±0.203	0.871±0.263	1.021±0.882
	L-SHADE	1.472±0.061	0.940±0.114	0.589±0.084	0.444±0.091	1.622±0.269	1.155±0.257	0.841±0.235	0.735±0.302
Glass identification	Original	4.627±0.379	3.010±0.165	2.277±0.080	1.726±0.089	4.589±1.377	4.128±0.929	3.535±1.138	2.922±1.010
	DE	4.147±0.598	3.009±0.165	2.260±0.074	1.710±0.064	5.284±1.061	4.137±0.926	3.539±0.876	2.851±1.017
	SHADE	4.015±0.188	2.424±0.188	1.532±0.348	1.114±0.298	4.815±1.413	4.676±0.300	3.853±1.454	4.495±2.115
	L-SHADE	4.092±0.175	2.527±0.356	1.547±0.247	1.132±0.274	5.393±1.705	4.595±1.164	4.034±1.747	3.974±1.089
Boston housing	Original	5.155±0.107	3.762±0.263	2.897±0.156	2.381±0.109	5.356±0.424	3.917±0.603	3.310±0.308	2.834±0.409
	DE	4.892±0.336	3.457±0.719	2.828±0.198	2.333±0.161	5.330±0.317	3.867±0.901	3.317±0.381	2.817±0.446
	SHADE	4.734±0.287	3.115±0.374	2.327±0.346	1.710±0.357	5.272±0.383	3.949±1.097	3.028±0.570	2.684±1.330
	L-SHADE	4.723±0.243	3.258±0.374	2.357±0.271	1.712±0.387	5.219±0.238	3.857±0.502	3.050±0.155	2.678±0.420
Red wine	Original	6.577±0.303	5.084±0.174	4.488±0.052	4.021±0.101	6.711±0.606	5.380±0.321	4.841±0.238	4.379±0.186
	DE	6.429±0.294	4.953±0.211	4.374±0.060	3.846±0.100	6.654±0.468	5.282±0.337	4.702±0.300	4.235±0.415
	SHADE	6.216±0.306	4.589±0.488	3.746±0.079	3.351±0.183	6.381±0.441	4.937±0.535	4.336±0.357	4.112±0.691
	L-SHADE	6.312±0.293	4.644±0.405	3.952±0.156	3.470±0.189	6.532±0.455	4.997±0.453	4.401±0.228	3.932±0.264
Parkinson	Original	7.325±0.244	5.488±0.225	4.256±0.130	3.492±0.102	9.121±1.297	7.574±1.371	6.019±1.066	5.403±0.501
	DE	6.800±0.254	5.013±0.712	3.431±0.488	2.993±1.054	8.718±0.602	6.903±1.369	6.554±1.537	5.179±0.555
	SHADE	6.472±0.430	4.544±0.631	3.162±0.144	2.279±0.351	8.909±1.506	6.825±2.117	7.410±4.073	6.781±1.995
	L-SHADE	6.638±0.250	4.635±0.572	3.114±0.336	2.296±0.311	8.622±1.259	7.103±1.726	6.476±1.706	6.011±2.681
Breast Cancer	Original	13.38±0.31	10.97±0.32	9.524±0.340	8.381±0.194	13.94±1.22	11.96±1.27	10.82±1.09	9.965±1.204
	DE	12.80±0.16	10.15±0.28	8.420±0.545	8.131±0.843	13.28±1.34	11.25±1.03	10.38±1.23	9.501±1.148

Wisconsin (Diagnostic)	SHADE	12.51±0.35	9.713±0.157	8.008±0.121	7.064±0.201	13.06±1.10	11.54±1.24	10.62±5.37	10.91±2.74
	L-SHADE	12.55±0.21	9.680±0.263	8.095±0.317	7.090±0.204	13.01±0.99	11.35±1.20	10.06±0.88	9.687±1.367
Wilt	Original	2.229±0.017	1.497±0.031	1.151±0.013	0.949±0.031	2.148±0.087	1.503±0.302	1.113±0.178	0.981±0.118
	DE	2.180±0.021	1.432±0.050	1.107±0.029	0.924±0.037	2.106±0.069	1.484±0.297	1.109±0.189	0.968±0.381
	SHADE	2.156±0.019	1.258±0.046	1.010±0.023	0.892±0.021	2.088±0.092	1.313±0.182	1.004±0.201	0.966±0.287
	L-SHADE	2.157±0.022	1.266±0.039	1.016±0.019	0.905±0.026	2.351±0.167	1.320±0.196	1.001±0.163	0.971±0.078
Combined Cycle Power Plant	Original	1.087±0.003	0.790±0.010	0.603±0.012	0.503±0.006	1.095±0.036	0.795±0.014	0.612±0.022	0.507±0.021
	DE	0.872±0.004	0.625±0.012	0.569±0.015	0.470±0.009	0.893±0.041	0.631±0.019	0.569±0.026	0.458±0.027
	SHADE	0.847±0.003	0.566±0.009	0.389±0.011	0.294±0.005	0.872±0.055	0.595±0.235	0.401±0.243	0.312±0.252
	L-SHADE	0.838±0.002	0.550±0.013	0.381±0.010	0.312±0.004	0.859±0.047	0.589±0.026	0.399±0.037	0.319±0.048
User Knowledge Modeling	Original	2.245±0.018	1.632±0.035	1.226±0.027	1.028±0.031	2.414±0.109	1.770±0.064	1.378±0.095	1.248±0.077
	DE	1.819±0.057	1.195±0.074	1.162±0.030	0.980±0.027	1.985±0.308	1.369±0.095	1.309±0.144	1.179±0.096
	SHADE	1.769±0.146	1.071±0.054	0.722±0.043	0.546±0.018	2.034±0.191	1.360±0.298	1.175±0.690	1.507±1.653
	L-SHADE	1.789±0.075	1.100±0.057	0.775±0.036	0.592±0.016	1.975±0.371	1.300±0.222	1.171±0.370	1.054±0.241
Seeds	Original	1.194±0.042	0.780±0.043	0.573±0.026	0.454±0.025	1.309±0.250	0.918±0.195	0.729±0.213	0.617±0.067
	DE	1.037±0.031	0.642±0.235	0.546±0.029	0.426±0.031	1.152±0.176	0.884±0.177	0.711±0.206	0.564±0.122
	SHADE	0.992±0.047	0.485±0.028	0.245±0.054	0.153±0.046	1.279±0.373	0.813±0.242	0.728±2.043	0.832±1.189
	L-SHADE	1.012±0.044	0.499±0.096	0.261±0.054	0.161±0.053	1.253±0.293	0.815±0.120	0.587±0.218	0.550±0.135
Gamma Telescope t	Original	4.314±0.016	3.401±0.024	2.817±0.023	2.526±0.019	4.336±0.170	3.431±0.171	2.847±0.083	2.561±0.098
	DE	4.125±0.027	3.154±0.035	2.685±0.046	2.215±0.050	4.132±0.243	3.156±0.249	2.718±0.254	2.539±0.194
	SHADE	4.038±0.023	2.986±0.019	2.110±0.037	1.921±0.034	4.281±0.251	3.319±0.313	2.702±0.140	2.367±0.205
	L-SHADE	4.022±0.019	2.997±0.024	2.091±0.020	1.957±0.026	4.095±0.210	3.102±0.215	2.198±0.179	2.116±0.157
Wholesale customers	Original	2.814±0.120	1.907±0.088	1.411±0.066	1.131±0.042	3.705±0.684	3.217±0.482	2.696±0.505	2.321±0.522
	DE	2.601±0.280	1.895±0.098	1.388±0.059	0.921±0.037	3.931±0.763	3.206±0.461	2.659±0.503	2.310±0.482
	SHADE	2.482±0.467	1.540±0.229	1.015±0.184	0.761±0.204	3.816±0.599	3.291±0.664	2.751±0.566	2.634±0.324
	L-SHADE	2.477±0.426	1.684±0.088	1.027±0.199	0.772±0.185	3.886±1.341	3.073±0.499	3.482±2.179	2.126±0.501

Table 3.3. Optimal values of the fuzzification coefficient *m* of the constructed models

	c = 6	c = 10	c = 15	c = 20
Auto-MPG	1.4±0.1	1.4±0.1	1.4±0.1	1.4±0.0
Glass identification	1.2±0.1	1.1±0.1	1.2±0.1	1.2±0.1
Boston housing	1.3±0.1	1.2±0.0	1.2±0.0	1.3±0.1
Red wine	1.1±0.0	1.2±0.1	1.2±0.0	1.2±0.0
Parkinson	1.3±0.0	1.3±0.1	1.3±0.1	1.3±0.1
Breast Cancer	1.2±0.0	1.2±0.0	1.2±0.0	1.2±0.0
Wilt	1.2±0.1	1.2±0.0	1.2±0.1	1.2±0.1
Combined Power	1.4±0.0	1.4±0.0	1.4±0.0	1.4±0.0
knowledge	1.3±0.0	1.3±0.0	1.3±0.0	1.3±0.0
seeds	1.4±0.0	1.4±0.1	1.5±0.1	1.5±0.0
Gamma Telescope	1.2±0.0	1.2±0.0	1.2±0.0	1.2±0.0
wholesale	1.2±0.1	1.2±0.0	1.3±0.1	1.3±0.1

Table 3.4. Statistical comparison of the augmented reconstruction scheme versus the original reconstruction approach

	Methods	Training data				Testing data			
		c = 6	c = 10	c = 15	c = 20	c = 6	c = 10	c = 15	c = 20
Auto-MPG	DE	+	+	+	+	+	o	+	+
	SHADE	+	+	+	+	+	+	o	o
	L-SHADE	+	+	+	+	+	+	+	o
Glass identification	DE	+	+	+	+	o	o	o	o
	SHADE	+	+	+	+	o	o	o	o
	L-SHADE	+	+	+	+	o	o	o	o
Boston housing	DE	+	+	o	o	o	+	-	+
	SHADE	+	+	+	+	o	o	o	o
	L-SHADE	+	+	+	+	o	o	+	+
Red wine	DE	+	+	+	+	+	+	+	+
	SHADE	+	+	+	+	+	+	o	o
	L-SHADE	+	+	+	+	+	+	+	+
Parkinson	DE	+	+	+	+	+	o	o	o
	SHADE	+	+	+	+	o	o	o	o
	L-SHADE	+	+	+	+	+	+	o	o
Breast Cancer Wisconsin	DE	+	+	+	+	+	+	o	+
	SHADE	+	+	+	+	+	o	o	o
	L-SHADE	+	+	+	+	+	+	o	o
Wilt	DE	+	+	+	+	o	o	o	o
	SHADE	+	+	+	+	o	+	o	o

	L-SHADE	+	+	+	+	+	+	+	+
Combined Cycle Power Plant	DE	+	+	+	+	o	+	+	+
	SHADE	+	+	+	+	+	o	o	o
	L-SHADE	+	+	+	+	+	+	+	+
User Knowledge Modeling	DE	+	+	+	+	+	+	+	+
	SHADE	+	+	+	+	+	+	o	o
	L-SHADE	+	+	+	+	+	+	o	+
Seeds	DE	+	+	o	+	+	+	+	+
	SHADE	+	+	+	+	o	o	o	o
	L-SHADE	+	+	+	+	+	+	+	+
Gamma Telescope	DE	+	+	+	+	+	+	o	o
	SHADE	+	+	+	+	o	o	o	o
	L-SHADE	+	+	+	+	+	+	+	+
Wholesale customers	DE	+	o	+	+	o	o	+	o
	SHADE	+	+	+	+	o	o	o	-
	L-SHADE	+	+	+	+	o	+	o	+

Table 3.5. Average optimization time (in hours) of the optimization algorithms

Methods	DE				SHADE				L-SHADE			
c	6	10	15	20	6	10	15	20	6	10	15	20
Auto-MPG	0.10	0.29	0.95	2.10	0.07	0.22	0.52	1.22	0.05	0.15	0.45	1.18
Glass identification	0.13	0.33	0.54	1.68	0.06	0.17	0.50	0.67	0.06	0.12	0.31	0.52
Boston housing	0.19	0.50	1.74	2.13	0.10	0.27	0.92	1.22	0.09	0.24	0.89	1.17
Red wine	0.40	1.11	3.13	4.94	0.24	0.71	1.70	4.12	0.21	0.64	1.64	4.03
Parkinson	0.24	0.56	1.48	3.04	0.13	0.32	0.82	1.61	0.10	0.23	0.53	1.54
Breast Cancer	0.75	1.83	3.51	6.84	0.29	0.76	1.94	2.12	0.24	0.56	1.67	2.01
Wilt	1.01	2.14	2.57	4.61	0.82	1.61	3.10	3.92	0.71	1.43	2.86	3.57
Combined Power	0.17	1.81	2.25	4.72	0.11	1.02	1.28	2.53	0.08	0.89	1.07	2.48
User Knowledge	0.08	0.26	0.87	1.99	0.05	0.16	0.48	1.07	0.05	0.12	0.31	0.64
Seeds	0.08	0.23	0.69	1.39	0.04	0.12	0.38	0.95	0.04	0.10	0.34	0.50
Gamma Telescope	5.12	9.73	12.4	21.6	2.16	4.83	9.61	16.2	1.94	4.15	8.71	15.8
Wholesale	0.10	0.45	1.27	2.67	0.06	0.17	0.61	1.30	0.05	0.12	0.58	1.17

The augmented reconstruction is based on the original reconstruction with the optimal values of the fuzzification coefficients m for 10-fold sets of each data. As reported in Table 3.5, the average optimal values of fuzzification coefficient and their standard deviations for these data sets are located far lower than 2. In particular, these values fall within the region ranging in-between 1.1 to 1.5, which is consistent with the conclusions made in [81].

In light of the reported experimental results, as shown in Table 3.3 and Table 3.5, the modified reconstruction method demonstrates a remarkably improved performance over the performance delivered by the original reconstruction approach. The reconstruction performances are enhanced significantly for almost all the training data and majority of the testing data. Furthermore, comparing the three optimization techniques, the performance of SHADE and L-SHADE has advantages over the training data set over the DE optimization, however if we need to make a trade-off on the performance over the training and testing data, the L-SHADE method is preferable. In addition, due to the execution time shown in Table 6, the L-SHADE method also has an advantage with regard to its computation efficiency. Thus, the L-SHADE method is particularly worth utilizing.

Moreover, as visualized in Table 3.5, the following parameters heavily impact the computing overhead: the size of the data (in terms of the number of data and their dimensionality) and the number of clusters. As to the performance of reconstruction, more clusters helps achieve better reconstruction results. One should be prudent with respect to the implementation of the optimization method. In some cases, when the number of clusters becomes higher (say $c = 20$), some overfitting occurs because of too many parameters to optimize. As a result, the interactive matrix and modified prototypes might be impacted by noise. In addition, since $D = c \times c + c \times n$, the number of optimized variables grows with the increasing number of clusters and the efficiency of the optimization might decrease. In sum, a lower number of clusters, say 5-15, is a sound alternative.

3.4 Summary

In this study, we have augmented the granulation-degranulation scheme to improve the performance of the reconstruction (degranulation) mechanism in the two different ways. The transformation of the originally developed partition matrices introduces both excitatory and inhibitory interaction between the membership grades of the individual data. The adjustment of the prototypes becomes helpful in the dynamic range expansion and compensating for the averaging effect reflective in the position of the prototypes being produced by the FCM method. Overall, the proposed method is beneficial to reduce the deterioration of the reconstruction results and enhance the performance of the overall granulation-degranulation scheme, which is meaningful for transforming data between numeric form and granular format. In addition, several population-based optimization algorithms have been used to carry out optimization. The comparative studies of different optimization methods convincingly demonstrate an importance of the selection of an efficient algorithm.

Chapter 4

Logic Operators and Granular Operations – A Statistical Review[b]

After obtaining the information granules, we embark on considering the process between the information granules. Logic operators, in the role of aggregation connectives, have been proven to be useful tools for information fusion. They are accordingly at the center of research into the theory of fuzzy sets and have subsequently fueled a vast array of applications. In granular computing, there is also a need for aggregating several information granules into a single output granule. Therefore, this chapter characterizes them by analytical properties.

The abundance of logic operations and aggregation operators is profound. In particular, the studies on triangular norms, nullnorms and uninorms [89]-[94] have been mushrooming. Analytical investigations have been blooming bringing a vast of theoretical findings. Interestingly, all the investigations were very much confined to analyses carried out at the *numeric* levels. Sometimes one may raise some arguments as to the applied relevance of developing so many t-norms and their eventual non-distinct behavior, which might benefit in a tangible way fuzzy reasoning schemes, fuzzy models, and fuzzy classifiers contributing to possible improvements of their performance. Notably almost all constructs of fuzzy sets use in one way or another fuzzy operators and composition operators, just to recall all schemes of approximate reasoning [95]-[98]. If we pose a bit and reflect upon the developments of fuzzy sets, we may eventually conclude that an over-emphasized *numeric* facet of fuzzy sets and their processing is somewhat counter-intuitive and does not lie in the spirit of any linguistic processing and system modeling, computing with words and alike – all these paradigms which were positioned at the heart of fuzzy sets from their very inception.

[b] A version of this chapter has been published as [129].

The problem we are concerned with is about determining whether different t-norms produce results that are indeed different and tangibly distinguishable. Being more precise, we check whether the results produced by two t-norms (or t-conorms) in the presence of some data ***D*** yield differences that are significantly different with the notion of difference being quantified in some well-defined statistical terms and being more precise articulated through testing appropriate statistical hypothesis.

Furthermore, generally, the logic operators deal with numeric aggregations of data or fuzzy sets information granules. But how about when the information granule is as the form of interval sets? Thus, here our line of thought is to explore another alternative. We propose a concept of granular operators implied by t-norms. Those operators are granular in the sense that we admit here a granular (interval-valued) outcome of the operation carried out on granular arguments. So that it "covers" the results produced by some other t-norms. In the realization of this granular construct, we engage the principles of Granular Computing and assess the quality of the granular operator in terms of its coverage and specificity criteria.

4.1 Logic Operators – A Brief Review

A t-norm is a binary operator T: $[0, 1]^2 \to [0, 1]$, defined in the unit square, which satisfies the following properties [90]:

(a) Commutativity: $T(x_1, x_2) = T(x_2, x_1)$

(b) Associativity: $T(x_1, T(x_2, x_3)) = T(T(x_1, x_2), x_3)$

(c) Monotonicity: $T(x_1, y) \leq T(x_2, y)$ if $x_1 \leq x_2$

(d) Boundary condition: $T(x_1, 1) = x_1$

A t-conorm S: $[0, 1]^2 \to [0, 1]$ can be defined as $S(x_1, x_2) = 1 - T(x_1, x_2)$. The first three properties of t-conorms are the same as those of t-norms while the boundary condition reads as $S(x_1, 0) = x_1$.

In the existing plethora of t-norms and t-conorms, we consider several representative families of t-norms and t-conorms, see Table 4.1. The t-norms reported here come with their duals.

Table 4.1. Selected examples of t-norms and t-conorms

Names	t-norms	t-conorms
Logical	$T_1(x_1,x_2)=\min(x_1,x_2)$	$S_1(x_1,x_2)=\max(x_1,x_2)$
Hamacher	$T_2(x_1,x_2)=\frac{x_1x_2}{x_1+x_2-x_1x_2}$	$S_2(x_1,x_2)=\frac{x_1+x_2-2x_1x_2}{1-x_1x_2}$
Algebraic	$T_3(x_1,x_2)=x_1x_2$	$S_3(x_1,x_2)=x_1+x_2-x_1x_2$
Einstein	$T_4(x_1,x_2)=\frac{x_1x_2}{1+(1-x_1)(1-x_2)}$	$S_4(x_1,x_2)=\frac{x_1+x_2}{1+x_1x_2}$
Lukasiewicz	$T_5(x_1,x_2)=\max(x_1+x_2-1,0)$	$S_5(x_1,x_2)=\min(x_1+x_2,1)$
Drastic	$T_6(x_1,x_2)=\begin{cases}x_1 & x_2=1\\ x_2 & x_1=1\\ 0 & x_1,x_2<1\end{cases}$	$S_6(x_1,x_2)=\begin{cases}x_1 & x_2=0\\ x_2 & x_1=0\\ 1 & x_1,x_2>0\end{cases}$
Triangular 1	$T_7(x_1,x_2)=\frac{2}{\pi}\cot^{-1}[\cot\frac{\pi x_1}{2}+\cot\frac{\pi x_2}{2}]$	$S_7(x_1,x_2)=\frac{2}{\pi}\tan^{-1}[\tan\frac{\pi x_1}{2}+\tan\frac{\pi x_2}{2}]$
Triangular 2	$T_8(x_1,x_2)=\frac{2}{\pi}\arcsin(\sin\frac{\pi x_1}{2}\sin\frac{\pi x_2}{2})$	$S_8(x_1,x_2)=\frac{2}{\pi}\arccos(\cos\frac{\pi x_1}{2}\cos\frac{\pi x_2}{2})$

4.2 Statistical Distinguishability of Logic Operators

Regarding statistical analyses of results produced by t-norms, there have been only a few studies reported so far, cf. [99][100]. The investigations covered here build upon the findings presented in the literature, however we move beyond them by looking at the family of triangular norm and the usage of data sets.

As a suitable test, we use here a nonparametric Mann–Whitney-Wilcoxon test [70][100]. It is more suitable than a well-known *t*-test [69] as it is more efficient on non-normal distributions than the *t*-test, and is nearly as efficient as the *t*-test for normal distributions [72]. The null hypothesis H_0 is formulated as follows

$$H_0\text{: medians of samples of results produced by } T_1 \text{ and } T_2 \text{ are equal} \quad (4.1)$$

Hence in testing hypothesis, it is essential to involve data of a certain nature. Two alternatives are considered with this regard:

Random data The data is drawn at random from a uniform distribution expressed over [0, 1].

Data produced through fuzzy clustering (FCM) Here we engage a clustering procedure to arrive at membership values and use them for testing purposes. There are two motivating factors here: clustering is commonly used to produce membership grades and clustering is quite commonly regarded as an essential design procedure in fuzzy modeling.

Once the data have been decided upon, either being drawn randomly or formed through the use of the FCM algorithm, the hypothesis H_0 is tested and conclusion formulated. In order to base an overall conclusion on a solid footing, the experiment is repeated a large number of times (say, 10,000) by involving different input data and a final count of the number of times when the hypothesis has been rejected is reported as well, and we compute the proportion r of statistically different tests as follows,

$$r = \frac{\text{number of times hypothesis} H_0 \text{ has been rejected}}{\text{number of data sets}} \tag{4.2}$$

The process described above (viz. produce the data set, test hypothesis, draw a conclusion) is carried out for both t-norms and t-conorms. In case of triangular norms with some parameters, the testing can involve different values of the parameters and checking how much distant the values should be to make the results statistically different.

4.3 Granular Indistinguishability of Triangular Norms

Instead of numeric arguments of T leading to the numeric outcome $T(x_1, x_2)$, let us assume that the arguments are granular, viz. intervals $X_1 = (x_1^-, x_1^+)$ and $X_2 = (x_2^-, x_2^+)$ having width 2ε and spanned over the original numeric values x and y

$$
\begin{aligned}
x_1^- &= \max(0, x_1 - \varepsilon) \\
x_1^+ &= \min(1, x_1 + \varepsilon) \\
x_2^- &= \max(0, x_2 - \varepsilon) \\
x_2^+ &= \min(1, x_2 + \varepsilon)
\end{aligned}
\tag{4.3}
$$

where ε assumes values ranging from 0 to 0.5, which is referred to as a level of information granularity [101][102]. Evidently, the result $Y = T(X_1, X_2)$ is also of interval form. In light of the monotonicity property of triangular norms, we have $Y = [y^-, y^+]$ where $y^- = T[x_1^-, x_2^-]$ and $y^+ = T[x_1^+, x_2^+]$. We can recall Y to be produced in this way a granular t-norm T (more specifically, it is a t-norm operating on granular-interval valued arguments). The assessment of indistinguishability of t-norms can be realized in the setting of Granular Computing. Given is a reference t-norm, call it T_0. Consider some other t-norm, T. We say that T and T_0 are ε–indistinguishable in light of data $\boldsymbol{D} = \{(x_1, x_2)\}$ if the following property is satisfied

$$
T(x_1, x_2) \in T_0(X_1, X_2) \text{ for all } (x_1, x_2) \in \boldsymbol{D} \tag{4.4}
$$

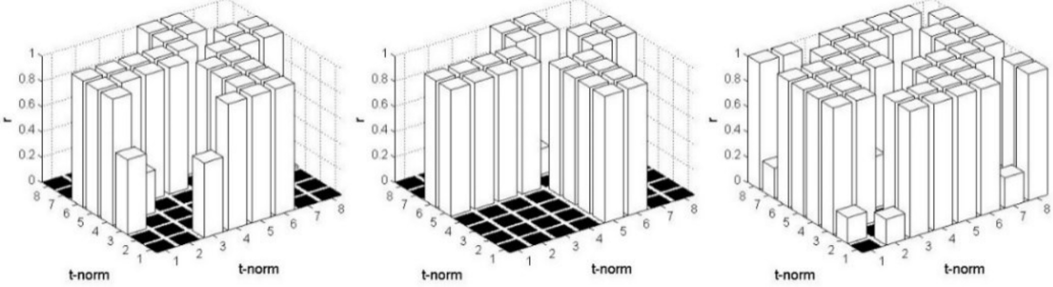
The property defined in this way is of binary character, namely the ε–indistinguishability holds or not. Obviously, its satisfaction depends upon the value of ε. Instead we can admit a gradual nature of its satisfaction by counting how many times the inclusion predicate $T(x_1, x_2) \in T_0(X_1, X_2)$ is true, incl($T(x_1, x_2)$, $T_0(X_1, X_2)$). Considering the data $\boldsymbol{D}$, we determine the coverage measure (normalized count of satisfaction of the inclusion predicate) as in (48). The plot of this measure versus ε can visualize the capabilities of making T indistinguishable from T_0. Thus, to produce some global descriptor of coverage, we determine area under the curve (AUC) as in (2.36), the higher the AUC value, the more indistinguishable T and T_0 are.

4.4 Experimental Studies

In this section, we present a suite of experimental results showing both the statistically-oriented and Granular Computing-based distinguishabilty properties of t-norms. Because of duality, we only report the comparisons involving t-norms and the results of t-conorms comparisons are the same. The level of significance is set as 0.05. Besides, the data produced through FCM clustering is Boston housing data set cited form UCI machine learning repository. The number of clusters (c) is selected as 2 and 10, and fuzzification coefficient (m) is set as 1.1 and 2.0.

4.4.1 Statistical Difference of T-norms

We test non-parametric t-norms listed in Table 4.1 and the results of r values are shown in Figure 4.1 Different t-norms (described their corresponding indexes, see Table 4.1) are positined on x and y-coordinates. In general, Lukasiewicz t-norm (T_5) and Drastic t-norm (T_6) behave significantly differently from other t-norms. The Algebraic (T_3) and Einstein (T_4) also demonstrate statistical difference in some cases, some of which were reported in the literature [99]. We can notice that, when the input data is coming from FCM clustering (c = 10, m = 2.0), most of the t-norms or t-conorms' outcomes are statistically different from the others. This also indicates that the inputs variables show an important effect on the values of r.



(a) (b) (c)

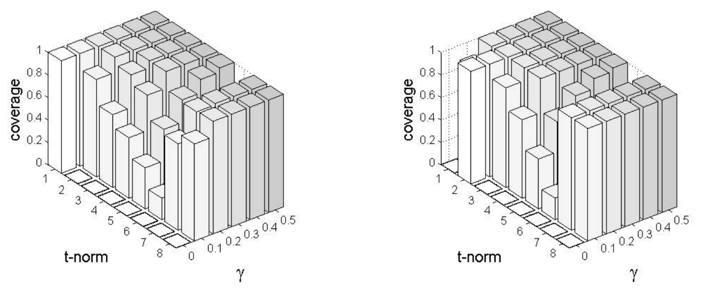
Figure 4.1. Values of r for eight non-parametric t-norms with data: (a) randomly distributed data, (b) results of FCM clustering (c = 2, m = 2.0), (c) results of FCM clustering (c = 10, m = 2.0)

4.4.2 Coverage Ratio of Granular Operators

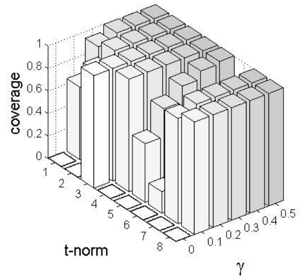
Figure 4.2 shows the coverage implied by every T_0 selection, and value of ε is set within the range of [0, 0.5]. X-axis shows the value of ε and Y-axis represents the subscript of T t-norms while Z-axis shows the coverage results. The inputs veriables exhibit a uniform random distribution. And Table 4.2 shows the *AUC* valuese when comparing the coverage performance.

As it can be seen in the bar plot of Figure 4.2, some T_0 selection have good coverage performance, such as Hamacher t-norm and Triangular t-norms, but some have worse coverage performance, such as Lukasiewicz t-norm and Drastic t-norm. Combining with the results we obtained above, we know that Lukasiewicz t-norm and Drastic t-norm are significantly different from other t-norms in most of the cases, so it can hardly cover or be covered by other granular operators, which is also testfied and shown in Figure 4.2 (e) and (f). Therefore, in the rest part of coverage computation, we exclude those two t-norms and evaluate the coverage performance of remaning 6 families of t-norms, as shown in Figure 4.2 (a) to (d) and Figure 4.2 (g) to (h). The inputs of those granular operators are 10,000 groups of pairs of data (x, y) uniformly distributed within the range of [0, 1], and we also tested groups of an equal amount of inputs obtained by FCM processing.

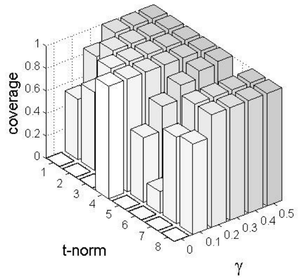
As shown in Figure 4.3, the coverages are monotonically ascending with extending the intervals γ of granularity. Those 6 sorts of t-norms approach to 100% at a certain value of γ. According to the results, we could conclude that T_0 selected as Hamacher t-norm (T_2) and Triangular t-norms (T_7, T_8) have better coverage performance, in other words, they can cover the results of granular operators with smaller granular intervals. In addition, the inputs also have some considerable impact on the increasing of coverage ratio. More specifically, for those six granular operators, the coverage performance for FCM clustering results is better than randomly distributed data sets.



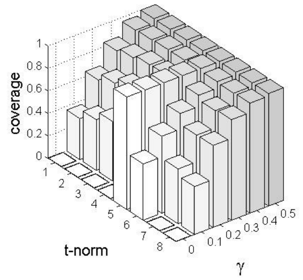
(a) (b)



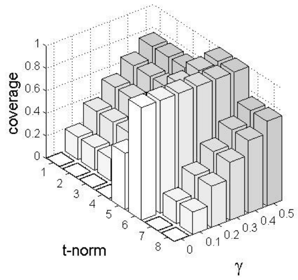
(c)



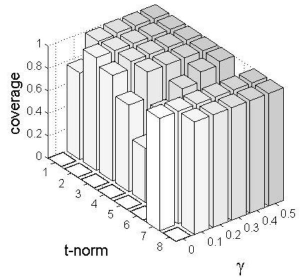
(d)



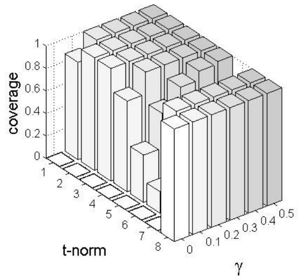
(e)



(f)

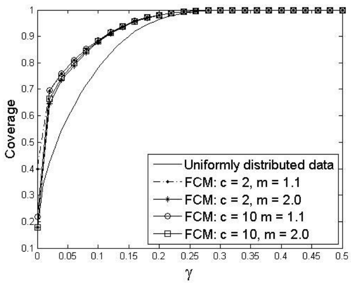


(g)

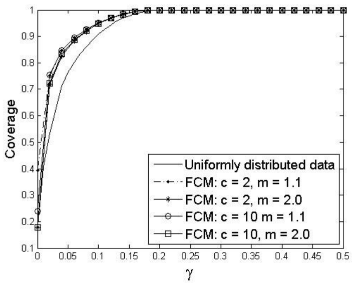


(h)

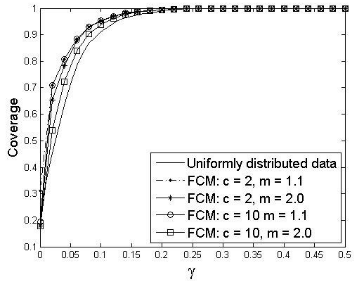
Figure 4.2. Coverage relationship obtained for granular operators with regard to other t-norms: (a) T_0 selected as T_1, (b) T_0 selected as T_2, (c) T_0 selected as T_3, (d) T_0 selected as T_4, (e) T_0 selected as T_5, (f) T_0 selected as T_6, (g) T_0 selected as T_7, (h) T_0 selected as T_8.



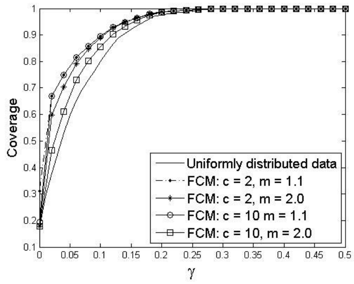
(a)



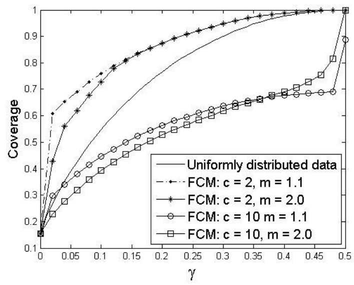
(b)



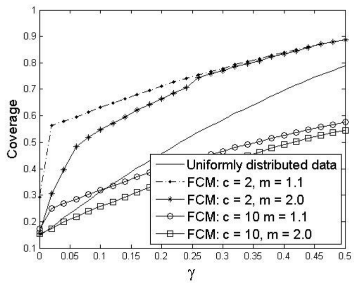
(c)



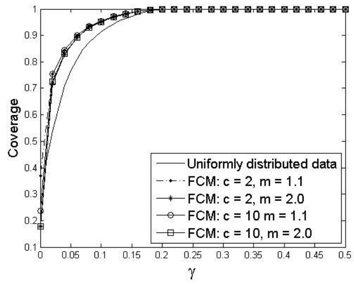
(d)



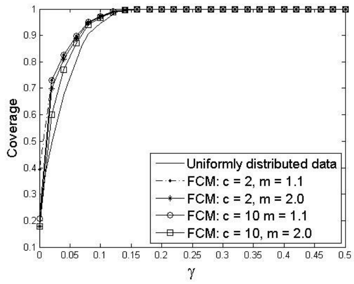
(e)



(f)



(g)



(h)

Figure 4.3. Coverage produced by different granular operators: (a) T_0 selected as T_1, (b) T_0 selected as T_2, (c) T_0 selected as T_3, (d) T_0 selected as T_4, (e) T_0 selected as T_5, (f) T_0 selected as T_6, (g) T_0 selected as T_7, (h) T_0 selected as T_8.

Table 4.2. *AUC* values obtained for comparing the coverage performance

Inputs	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
Random data	0.4503	0.4725	0.4681	0.4518	0.3877	0.2572	0.4730	0.4741
FCM: c = 2, m = 1.1	0.4770	0.4885	0.4853	0.4766	0.4409	0.3728	0.4885	0.4893
FCM: c = 2, m = 2.0	0.4718	0.4845	0.4817	0.4716	0.4307	0.3459	0.4847	0.4855
FCM: c = 10, m = 1.1	0.4752	0.4870	0.4842	0.4755	0.2870	0.2154	0.4872	0.4875
FCM: c = 10, m = 2.0	0.4730	0.4848	0.4757	0.4624	0.2811	0.1908	0.4850	0.4815

4.5 Summary

In this chapter, we have offered a qualitative view at logic operators realized by triangular norms by comparing them in terms of their statistical non-distinguishability. The alternative approach is formulated in terms of information granules involving granular logic operators. In both ways, logic operators are cast in a more abstract setting thus supporting a new direction in more abstract way of treating and interpreting fuzzy computations and fuzzy modeling. These investigations are beneficial in identifying distinct groups of t-norms thus facilitating emergence substantial differences among results produced by t-norms.

Chapter 5

Granular Input Space of Fuzzy Rule-Based Models[c]

From this chapter, we start to develop granular fuzzy rule-based models based on the framework of numeric fuzzy rule-based models. Commonly, when we take advantage of a model, the first thing we can manipulate is the input. The granular input space plays a vital role. Here our focus is on the construction of inputs of the model in the form of information granules with the objective of gaining better insight into the roles the individual input variables play in the model. The investigations dealing with the exploration of the granular input space and its construction are subjects of this study.

It becomes obvious that different input variables exhibit a different impact on the outputs of the model. A quantification of this impact relates to some sensitivity analysis. Here we formulate the problem in the setting of information granules. We are interested in the following problem: if we make some input variable to become an information granule, what is its impact on the granularity of the result? Assuming that we have at our disposal a certain level of information granularity (regarded as a useful design asset), the problem is: how this asset becomes distributed (allocated) across all input variables so that the granular output exhibits the highest level of information granularity as possible? In this setting information granules and their level of information granularity (specificity) are important design characteristics to be optimized. Intuitively, one may anticipate that in this allocation exercise, an input variable whose impact on the output is quite limited, comes as an information granule of low specificity. The opposite holds true in case of any input variable, which significantly exhibit the output of the model. Here to retain a high specificity of the output results, this input variable has to be kept quite specific. It is important to note that information granules associated with the certain input variable is a tangible, easily interpretable and practically sound outcome of the process of allocation of information

[c] A version of this chapter has been published as [130].

granularity. Consider that we have some limited resources to acquire input data and an acquisition of the input variables comes with a certain cost. The precision of the acquired input variable is related with the granularity of the variable: higher granularity implies higher precision (specificity) of the variable. Our intent is to distribute the resources so that the quality of the granularity of the result produced becomes maximized, viz. its specificity is the highest one.

The objective of this study is to establish a systematic way of allocation information to input variables by solving a certain optimization problem of specificity maximization. An overall view of the optimization framework is schematically visualized in Figure 5.1. We proceed with an existing fuzzy model (which is a numeric mapping $y = f(\boldsymbol{x})$) and establish a way of allocating information granularity across the input variables (making them granular (denoted here by $\boldsymbol{X}$) so that the granular output of the fuzzy model $Y = f(\boldsymbol{X})$ exhibits the highest level of specificity. In essence, the optimization problem boils down to a development of a *granular* input space.

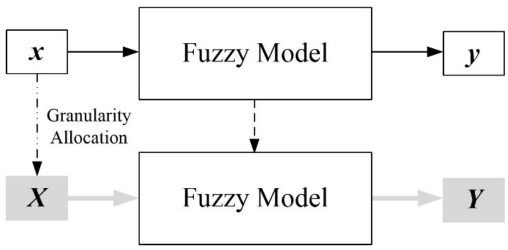


Figure 5.1. The concept of granular input space in fuzzy model systems

5.1 Formation of Granular Inputs of Fuzzy Models

The specificity measures the quality of the outputs being specific rather than general. In a concise and formal way, we express the problem in the following way.

Given some input data and a certain predetermined level of information granularity ε, realize an optimal allocation of this information granularity across the input variables so that the (5.1) *specificity of the corresponding output of the fuzzy model is maximized.*

In what follows, we elaborate on the related details that bring more clarification to the problem. The vector of numeric inputs $\boldsymbol{x}$ is made granular, denoted here by $\boldsymbol{X}$, viz. the coordinates of the vector become information granules, say intervals. The granular inputs $\boldsymbol{X}$ are formed by allocating levels of information granularity to the individual coordinates of the input $\boldsymbol{x}$ so that the requirement $\varepsilon_1^+ + \varepsilon_1^- + \varepsilon_2^+ + \varepsilon_2^- + \ldots + \varepsilon_n^+ + \varepsilon_n^- = n\varepsilon$ is satisfied where ε_1^+, ε_1^-, ε_2^+, ε_2^-, etc. assume values in the unit interval, ε_i^+, $\varepsilon_i^- \in [0,1]$, associated with the corresponding numeric values $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. We form the granular (interval-valued) input $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_t, \ldots, \boldsymbol{X}_n]$ coming in the following way

$$\boldsymbol{X}_t = [\boldsymbol{x}_t^-, \boldsymbol{x}_t^+] = [\boldsymbol{x}_t - \varepsilon_t^- range_t, \boldsymbol{x}_t + \varepsilon_t^+ range_t] \tag{5.2}$$

$t = 1, 2, \ldots, n$. The range of the corresponding input is expressed in a standard way

$$range_i = \max(x_t) - \min(x_t) \tag{5.3}$$

The way of forming the interval-valued input can be interpreted as follows: we make it interval-valued with the interval positioned symmetrically around the original input value with the width being a fraction of the range of the corresponding variable. It is apparent that the higher the value of ε_i, the broader the interval $\boldsymbol{X}_i$.

The specificity of the result generated by the granular model is response to the input $\boldsymbol{X}$ (which is some interval $Y = [y^-, y^+]$) is regarded as in (5.2), and the optimization problem can be expressed as the maximization of the specificity under the formula (2.35), which is subjected to constrains ε_i^-, $\varepsilon_i^+ \in [0, 1]$ and $\sum_{i=1}^{n} \varepsilon_i^- + \sum_{i=1}^{n} \varepsilon_i^+ = n\varepsilon$.

One has to remark here that the optimal allocation of granularity is data dependent. The calculations of the granular output of the rule-based fuzzy model require attention. For a given $\boldsymbol{X}_k$, we determine its membership grade to each information granule (being described by the corresponding prototype) by determining the following bounds,

$$\alpha_{k,i}^{1} = \frac{1}{\sum_{j=1}^{c} \left(\frac{|| \boldsymbol{x}_k^- - \boldsymbol{v}_i ||}{|| \boldsymbol{x}_k^- - \boldsymbol{v}_j ||} \right)^{2/(m-1)}} \tag{5.4}$$

$$\alpha_{k,i}^{2} = \frac{1}{\sum_{j=1}^{c} \left(\frac{|| \boldsymbol{x}_k^+ - \boldsymbol{v}_i ||}{|| \boldsymbol{x}_k^+ - \boldsymbol{v}_j ||} \right)^{2/(m-1)}} \tag{5.5}$$

Then the degree of membership (activation level) of the *i*-th rule becomes an interval $[\alpha_{k,i}^-, \alpha_{k,i}^+]$ with the bounds expressed as follows,

$$\alpha_{k,i}^- = \min\left(\alpha_{k,i}^1, \alpha_{k,i}^2\right) \tag{5.6}$$

$$\alpha_{k,i}^+ = \max\left(\alpha_{k,i}^1, \alpha_{k,i}^2\right) \tag{5.7}$$

In the sequel, the interval-valued output of is computed in the form

$$\left[y_k^-, y_k^+\right] = \sum_{\substack{i=1 \\ \oplus}}^{c} \left[\alpha_{ki}^-, \alpha_{ki}^-\right] \otimes \left[f_i^-(\boldsymbol{X}_k), f_i^+(\boldsymbol{X}_k)\right] \tag{5.8}$$

where

$$\left[f_i^-(\boldsymbol{X}_k), f_i^+(\boldsymbol{X}_k)\right] = \left[\omega_i, \omega_i\right] \oplus \left[\boldsymbol{a}_i^T, \boldsymbol{a}_i^T\right] \otimes \left[\boldsymbol{x}_k^- - \boldsymbol{v}_i,\ \boldsymbol{x}_k^+ - \boldsymbol{v}_i\right] \tag{5.9}$$

The operations shown in circles indicate that the operations of addition and multiplication are carried out for intervals rather than numbers as the calculations of interval arithmetic.

The optimization problem (5.1) is challenging given the complex dependence of the specificity of the output vis-à-vis the individual levels of information granularity. We consider a variant particle swarm optimizer (PSO) approach, comprehensive learning

PSO (CLPSO) [103] is considered. Once the optimization has been completed, the level of granularity associated with the *t*-th input variable can be described as the following sum

$$\ell_t = \varepsilon_t^- + \varepsilon_t^+ \tag{5.10}$$

$t = 1, 2, \ldots, n$. To make this sum independent of the level of information granularity ε, we consider the integral I_t

$$I_t = \int_0^{\varepsilon_{max}} \ell_t(\varepsilon) d\varepsilon \tag{5.11}$$

where $\varepsilon_{\max}$ is the maximal value of ε used in the allocation of information granularity, $\varepsilon_{\max} < 1$.

5.2 Global Sensitivity Analysis

In its generic way, a global sensitivity analysis is aimed at the exploration of impact of the variance of outcomes of a certain model attributed to the variability of input variables. As such, this analysis is commonly exploited as a sensitivity analysis along with other approaches including regression method, elementary effect method, meta-model based method and variance-based method [104]-[106]. The variance-based global sensitivity analysis [105]-[107] is concerned with the relationships assuming the following format

$$y = g(\boldsymbol{x}) \quad g: R^n \to R \tag{5.12}$$

where y is the output of the model, $\boldsymbol{x}$ is the vector of n variables, and g is the model function. The underlying idea is to quantify the variance of the output implied by the variance of different input variables and express an interaction present among the variables. Typically, such calculations rely on the Sobol's sensitivity indexes

[106][108]. They describe an impact of the individual variable, pairs of input variables, triples of them, etc. by invoking the following formulas

$$S_i = \frac{TV_i}{TV(y)}, \quad S_{ij} = \frac{TV_{ij}}{TV(y)} - S_i - S_j, \quad S_{ijk} = \cdots \tag{5.13}$$

$$S_{T_i} = S_i + S_{ij} + S_{ijk} + \cdots = \sum S_{i\sim} \tag{5.14}$$

where $TV(y)$ is the total variance of the output y, TV_i is the variance of the output being affected by the i-th input variable, and TV_{ij} is the variance of the output being impacted by the i-th and j-th input variable. In other words, the so-called first order index S_i quantifies the contribution of a single individual input variable to the variance of the output. The second order index S_{ij} expresses the contribution of interaction of the i-th and j-th input variables, and so on for higher order effects ($S_{i\sim}$, which represents the effects indices containing the interaction influence of the i-th input variable along with any other variables). Total order effect S_{Ti} is used to express the overall output sensitivity variance influenced by the i-th input variable, including all variance caused by its interaction with any other input variables. The larger value of the index becomes, the greater impact the input variable exhibits on the output.

5.3 Experiment Studies

We report a series of experiments here concerns some publicly available data sets coming from UCI machine learning repository and CMU StatLib library (http://lib.stat.cmu.edu/datasets/) to investigate the proposed approach. we carried out FCM-based clustering for different number of clusters (c) and values of the fuzzification coefficient (m) to develop the fuzzy model. In the sequel, the inputs were made granular realizing a process of optimal allocation of information granularity. Each data is split into 10-fold, and for each predetermined level of information granularity ε the experiment was repeated 10 times.

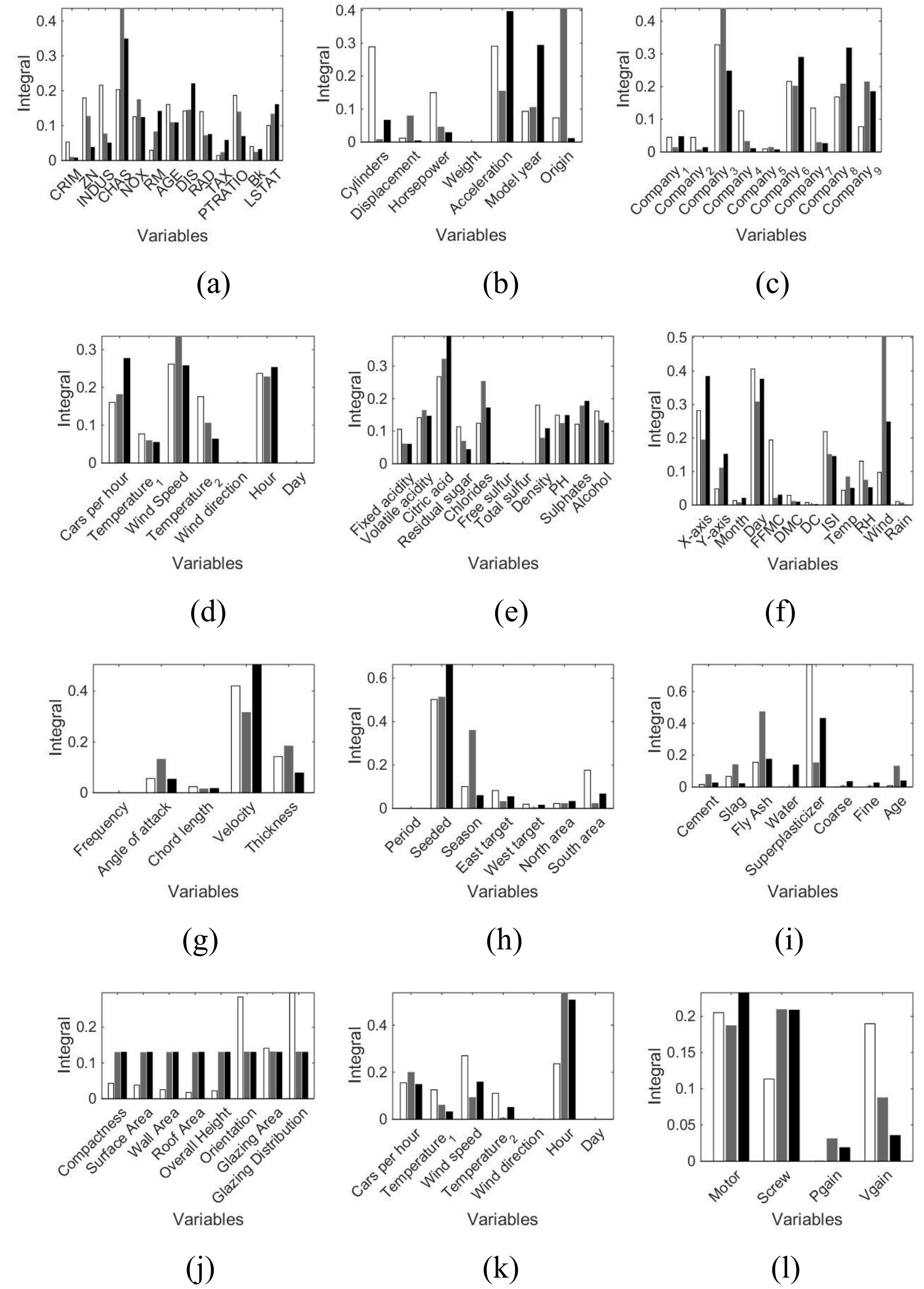


Figure 5.2. The integral of optimal granularity interval of the fuzzy models (m = 2.0): (a) Boston housing, (b) Auto MPG, (c) Stock, (d) PM10, (e) Red wine, (f) Forest fires, (g) Airfoil Self-Noise, (h) Cloud, (i) Concrete Strength, (j) Energy efficiency, (k) NO2, (l) Servo. The white bars: c = 2, the gray bars: c = 5, and the black bars: c = 9.

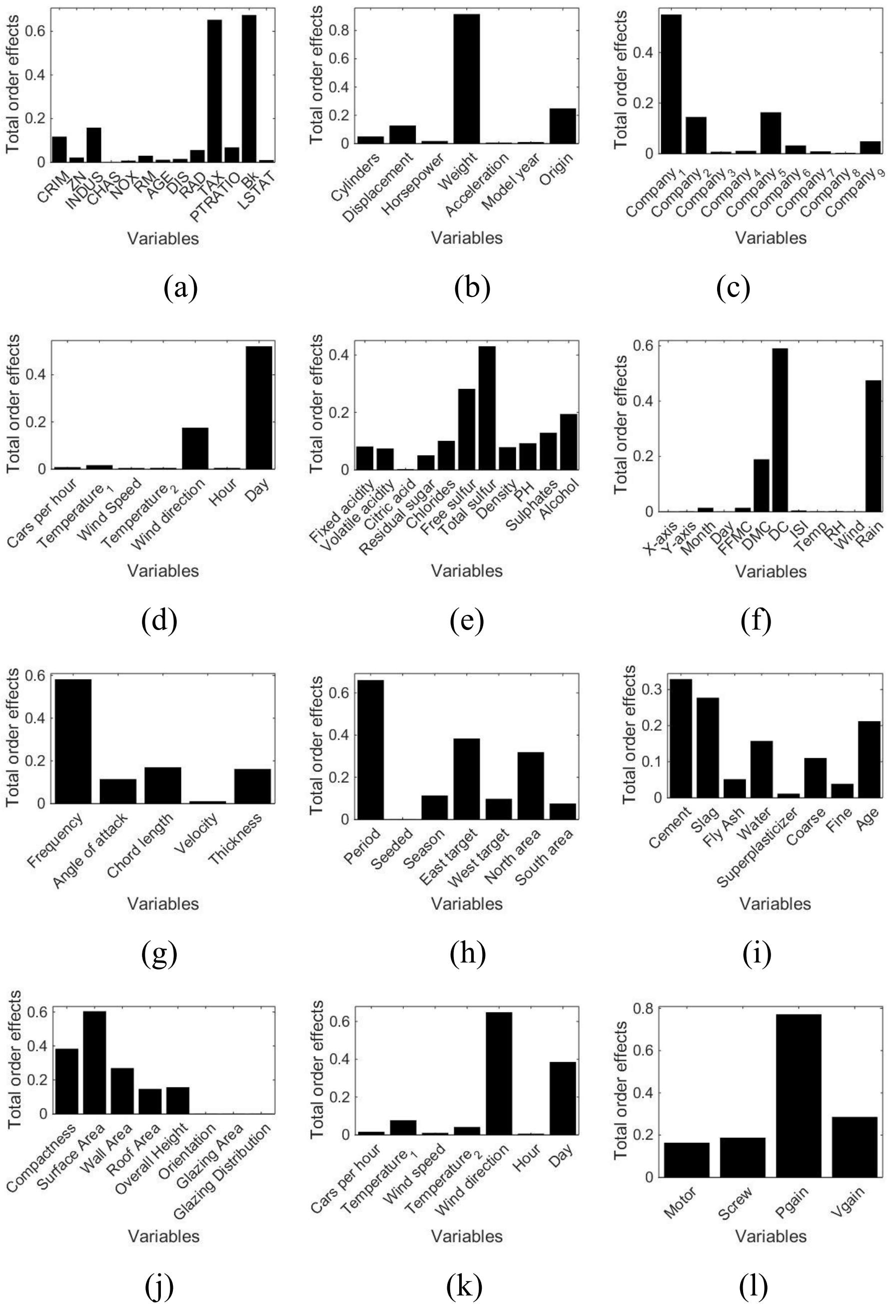


Figure 5.3. Results of global sensitivity: (a) Boston housing, (b) Auto MPG, (c) Stock, (d) PM10, (e) Red wine, (f) Forest fires, (g) Airfoil Self-Noise, (h) Cloud, (i) Concrete Strength, (j) Energy efficiency, (k) NO2, (l) Servo.

Figure 5.2 and Figure 5.3 summarize the results of the integral of optimal granularity interval of fuzzy models and the global sensitivity analysis results for the corresponding data sets.

A number of low sensitive input variables are associated with broader intervals. As an example, consider the Airfoil Self-Noise data. The frequency input exhibits higher sensitivity, see Figure 5.3 (g), thus narrow intervals are allocated to that variable, Figure 5.2 (g). In contrast, velocity variable has the lowest value of sensitivity and as such comes with the highest value of the integral of granularity allocation. In conclusion, resorting to CLPSO approach, the granularities tend to allocate to the low sensitive input variables to void high influence on the outputs, and vice versa, so that it is helpful to improve the specificity of outputs. Nevertheless, it should be noticed that, the proposed method is used for modeling purpose and allows us to use the data in various resolutions efficiently. Though some significant characteristics of the results obtained by the proposed approach and the standard sensitivity analysis are verified mutually, they cannot be substituted by each other. Because, when the sensitivities of two variables are very close, it is difficult for granularity allocation to distinguish them from each other. Meanwhile, sensitivity is not the only condition of input variables granularity allocation.

5.4 Summary

In this study, we have focused on the concept of the granular input space and its optimization. The direct linkages between the granular space formed in this way and the analysis of impact of input variables in the already developed models have been identified. The granular input space delivers an interesting vehicle to realize a global sensitivity analysis and offer a way of forming information granules of input variables.

Chapter 6

Granular Processing Module of Fuzzy Rule-Based Models

In this chapter, we study several different ways to design the granular fuzzy rule-based models by generalizing the processing modules and parameters of fuzzy rule-based models to granular forms.

6.1 Granular Interactive Rule Matrix of Fuzzy Models[d]

Initially, we are interested in the fuzzy rules, which contain prior knowledge obtained from learning processes or experts. Our focus is on two questions in this study: How are the rules generated, and what is the relationship between the rules? A spectrum of studies appears in the literature that are focused on the analysis of fuzzy models and that elaborate on the knowledge representation realized by such modules. There are also discussions reported on the quality of the rules and investigations of various reasoning or aggregation mechanisms [109]-[112]. Possible interactive relationships between the rules are usually discussed and realized by means of mechanisms of aggregation or decomposition, and subsequently implemented to improve the performance of fuzzy models, fuzzy classifiers and decision-making schemes [113]-[116]. The concept of formation and interaction within hierarchies of fuzzy models is related to some aspects of deep learning, in which we inherently encounter a collection of layers of processing facilitating a formation of additional features emerging at the higher levels of abstraction[117][118]. Likewise, the available design procedures are also highly diversified both in terms of general development strategies as well as the detailed optimization vehicles being used in their construction [114][115].

Interestingly, within the plethora of architectural enhancements of fuzzy rule-based models, a question of interaction among the rules and a possible usage of the interaction mechanism towards the enhancements of the performance of the rule-based models has

[d] A version of this chapter has been published as [131].

not been studied. By interaction among the rules we mean a mechanism of engaging existing fuzzy sets standing the corresponding rules in a formation of new fuzzy sets so that the fuzzy sets formed in this manner give rise to more efficient structure of the overall model. This is an original and promising development direction whose main features and associated improvements are worth considering. One objective in this study is to introduce the concept of interactive rules, study their properties and show ways they deliver additional functionality to fuzzy models. Afterwards, we augment the numeric model by building its granular augmentation (generalization). Here, information granularity is introduced into the already formed rule-based model by making the weights of the interaction matrix granular.

6.1.1 Forming Interaction Among Fuzzy Rules

A TS architecture of the fuzzy model is considered, and a generic architecture of the model dwells upon a collection of "*c*" rules assuming in the form as (2.6). Here we focus on the interaction between the fuzzy rules, so $f_i(\boldsymbol{x})$ is simply regarded as some numeric representatives y_i distributed across the output space. The output of inference (reasoning) scheme is as shown below

$$\hat{y} = \sum_{i=1}^{c} A_i(\boldsymbol{x}) y_i \tag{6.1}$$

In general, the design of rule-based models entails two fundamental phases:

(i) *granulation of input space*. The input space is granulated by forming a collection of fuzzy sets ($\boldsymbol{A}_i$) and in this way revealing a structure in the space of input data. Commonly FCM is used here. From the design perspective, the main parameters of the FCM such as the number of clusters c (number the rules) and the fuzzification coefficient m can be adjusted according to the needs of fuzzy modeling.

(ii) *determination of the conclusions of the rules*. This design phase follows the formation of information granules in $\boldsymbol{R}^n$ and is concerned with a specification of a class of local models (type of functions f_i) and estimation their parameters. Given that the type of the function has been already fixed (m_i), the values of these constants are optimized in a supervised mode by minimizing a certain performance index Q in such a way that a distance between the data and the corresponding results produced by the model are made as close as possible. The supervised mode of learning (estimation) is based in input-output data ($\boldsymbol{x}_k$, $target_k$) and quite commonly MSE is expressed as sum of squared errors

$$MSE = \sum_{k=1}^{N} \left(target_k - \hat{y}_k\right)^2 \tag{6.2}$$

where $\hat{y}_k$ is the output of the rule-based model produced for the given input $\boldsymbol{x}_k$; y_k= FM($\boldsymbol{x}_k$), where FM is governed by (6.2). Likewise, one can consider an $RMSE$ version of (6.3) taking on the following form

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(target_k - y_k)^2} \tag{6.3}$$

For the design purposes, the data are split into training and testing parts or eventually training-validation-testing parts and the corresponding values of $RMSE$ index serve as indicators of the performance of the model.

There are discussions reported on the quality of the rules and investigations of various reasoning or aggregation mechanisms [109]-[116]. Here, interactive rules are augmented by bringing some efficient schemes so that the introduced effect of interaction can be exploited to improve the quality of the resulting model. In general, the interaction can be accomplished in a form of a certain mapping g transforming original fuzzy sets standing in the rules, namely A_1, A_2, …, A_c into new fuzzy sets A^*_1, A^*_2, …, $A^*_{c'}$ where each newly formed fuzzy set is based upon the original fuzzy set. In an explicit way, we can consider A^*_i to be formed with the aid of a certain interaction matrix W followed by a nonlinear mapping g, viz.

$$A^*_i = g(A_1, A_2, \ldots, A_c) \tag{6.4}$$

$i = 1, 2, \ldots, c'$. Formally, the above mapping g realizes a transformation from $[0, 1]^c$ to [0,1]. From the point of view of the processing realized in this way, we encounter a new processing module of the architecture of the rule-based model as displayed in Figure 6.1.

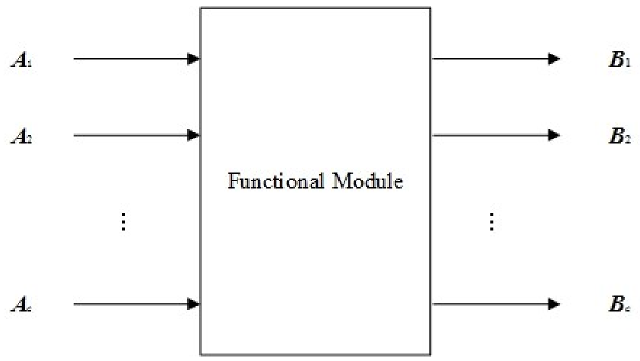


Figure 6.1. Augmentation of rule-based models by a functional module realizing interaction among the rules

We consider a class of mappings (transformations) in the concise form

$$A^*_i = g(WA) \tag{6.5}$$

W is a matrix of interactions whose usage supports a transformation of the fuzzy sets $A_1, A_2, .., A_c$ into the new transformed fuzzy sets. The entries of the matrix assume numeric values coming from some predetermined range of values being distributed symmetrically around 0 and embracing both negative and positive values. This helps realize both inhibitory as well as excitatory influence exerted by the original fuzzy sets. From the structural point of view, we note that the augmented fuzzy models exhibit an additional processing layer and help realize some additional faculties of deep learning.

Expressing explicitly the fuzzy sets, we obtain the following expression

$$y = \sum_{i=1}^{c'} \left[g\left(\sum_{j=1}^{c} W_{ij} \boldsymbol{A}_j(\boldsymbol{x}) \right) \right] \bar{y}_i \tag{6.6}$$

With regard to the nature of the interaction completed in terms of (64), we distinguish three cases:

(a) the same number of new fuzzy sets, namely $c = c'$. The new fuzzy sets come in the same number however they arise through an interaction among A_i.

(b) reduction effect, $c' < c$. Through this transformation, we end up with the lower number of fuzzy sets and effectively, the lower number of rules

(c) expansion effect, $c' > c$. Here we form a larger collection of fuzzy sets and in this way, increase the number of rules.

The interaction among the rules, which is realized through the mapping presented above (64) is quantified by taking a sum of absolute values of entries of W outside the main diagonal.

6.1.2 Realization of Interaction Fuzzy Rule-Based Models

Proceeding with the detailed discussion of the augmented architectures of the rule-based models, we consider several optimization strategies.

topology-1 This model serves as a reference structure. Its formal description is given by (40).

topology-2 here we have an augmented version of the reference model where the output is described in the form by (6.6). The nonlinear function g is specified as a limiter taking on the form

$$g(\theta) = \begin{cases} \theta & \text{if } \theta \in [0,1] \\ 1 & \text{if } \theta > 1 \\ 0 & \text{if } \theta < 0 \end{cases} \tag{6.7}$$

The entries of the transformation matrix are confined to the [-2, 2] interval.

topology-3 this model is described by (6.6) but now the nonlinear function is specified to be a sigmoidal one, namely

$$g(\theta)=\frac{1}{1+\exp(-\alpha\theta)} \tag{6.8}$$

where α (>0) is a steepness factor of the sigmoidal function.

Following these topologies of the models, we list several development scenarios, which along with the topologies of the models lead to five models of rule-based architectures:

Model-1 here we optimize the conclusions of the rules, namely $y_1, y_2, \ldots, y_c$.

Model-2 as the structure is augmented by the transformation matrix W, its entries are now optimized, and assume that $c = c'$.

Model-3 in this structure, in addition to learning W, the steepness factor is also optimized. The observation as made in case of *Model-2* holds here.

Model-4 we make the development as discussed in *Model-2* more advanced by allowing learning $\boldsymbol{W}$ as well as $y_1, y_2, \ldots, y_c$.

Model-5 here we follow *Model-3* but we optimize the constant values in the conclusion parts $y_1, y_2, \ldots, y_c$ of the rules.

6.1.3 Allocation of Granular Interaction Fuzzy Rule-Based Model

Information granularity is introduced into the already formed rule-based model by making the weights of the interaction matrix granular. These granular weights can be built around the original numeric entries of the weight matrix, W_{ij}, i, j =1, 2,…, c in many different ways and the strategy itself can be optimized. Here we consider two optimization scenarios in which information granules are constructed for the numeric parameters as shown in Figure 6.2.

(I) The first realizes a symmetric and uniform allocation of the level of information granularity. Given some predefined level of information granularity ε, we compute the product $\varepsilon \cdot range$ (where range is the length of the values the weights W_{ij} assume, say for the values of the weights ranging from -1 to 1, meaning that the *range* is 2) and build a symmetric interval around W_{ij} with the bounds expressed as follows,

$$\left[W_{ij} - \frac{\varepsilon}{2} range, W_{ij} + \frac{\varepsilon}{2} range\right] \quad (6.9)$$

where if necessary, the bounds of the above interval are clipped so that they are retained within the interval [-2, 2].

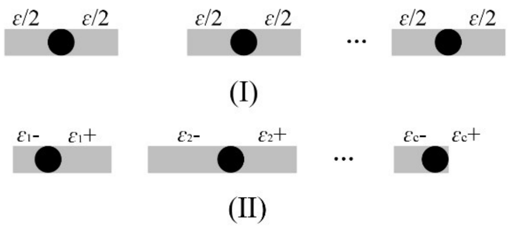


Figure 6.2. Two scenarios of allocation of information granularity: uniform (I) and nonuniform (II)

(II) We form intervals that exhibit a uniform allocation (are of the same length) but are asymmetrically distributed around the numeric value. The resulting interval comes in the form

$$[W_{ij} - \gamma \cdot \varepsilon \cdot range \,, \; W_{ij} + (1-\gamma) \cdot \varepsilon \cdot range] \quad (6.10)$$

where γ is a so-called asymmetricity index assuming values in the unit interval. The values of γ can be subject to optimization.

The criteria used to assess the performance of the granular model concerns coverage of the data and is taken as (2.29). we also compute the specificity of the interval produced by the granular model as (2.32). The criterion used to assess the performance of the granular model concerns both coverage and specificity, so the global indicator used here is as in (2.36).

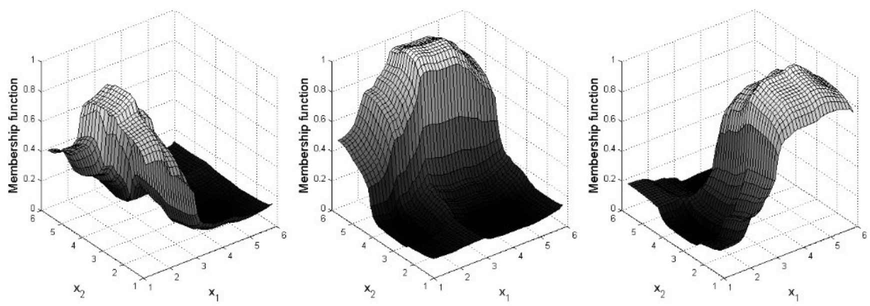
6.1.4 Experimental Studies

we present a synthetic data experiment to demonstrate the performance of the proposed method. A low-dimensional synthetic data is used here for illustrative purposes. We are concerned with the two-variable nonlinear function [65][67] described in the following form

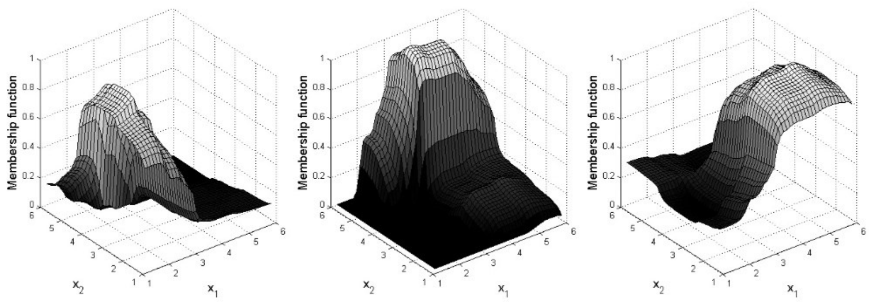
$$y = (1 + x_1^{-0.5} + x_2^{-1})^2 \tag{6.11}$$

where the two independent variables (x_1 and x_2) assume values in [1, 6]. The learning is realized by using a data set composed of 676 data pairs where the inputs are distributed randomly in the Cartesian product [1, 6]×[1, 6]. The 10-fold cross validation method is used. In the experiments, the range of the entries of the weight matrix W_{ij} is restricted to the interval [-2, 2]. The range of the parameter α standing in the sigmoidal function is confined to [0.1, 30], which is sufficient to make the function exhibiting a broad spectrum of steepness levels.

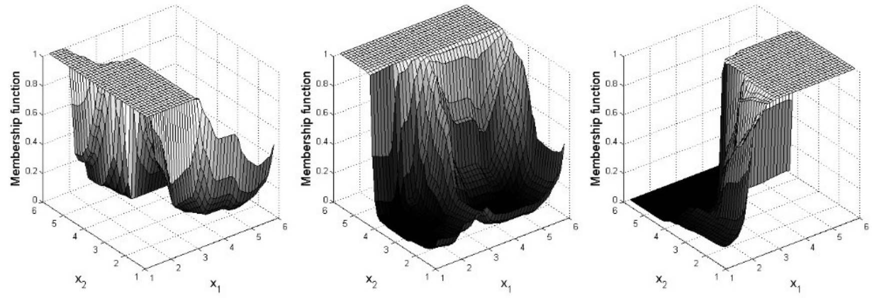
To contrast the behavior of the generic model and its successive refinements, some example membership functions of A_i and those obtained after their nonlinear transformation are displayed in Figure 6.3. It is noticed that the original membership functions have been affected because of the optimized interaction among them.



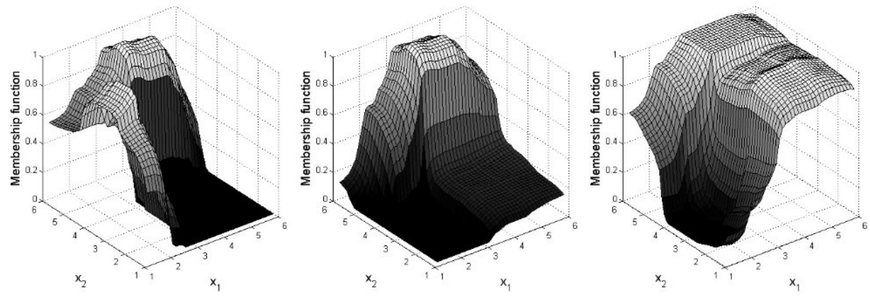
(a)



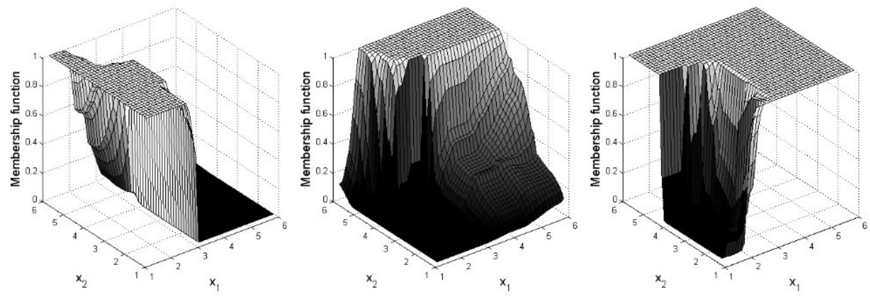
(b)



(c)



(d)



(e)

Figure 6.3. Plots of membership functions of Ai obtained for the fuzzy models for c = 3 and m = 2: (a) model-1, (b) model-2, (c) model-3, (d) model-4, (e) model-5.

Proceeding with a more comprehensive experimentation we developed the models for selected values of the number of rules (c) and the values of the fuzzification coefficient m set to 1.1, 2.0, and 2.8. The range of admissible values of α was set to [0, 30]. The results are reported by displaying the values of the performance index obtained by running the 10-fold cross validation; both the mean values and the corresponding standard deviations are presented, refer to Figure 6.4.

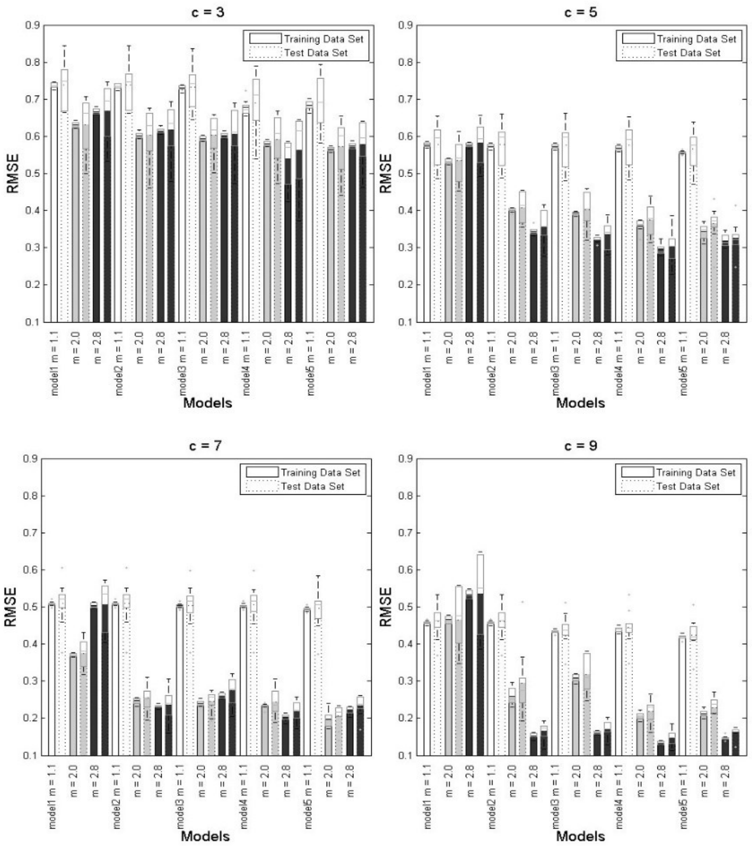


Figure 6.4. Performance of the models obtained for selected values of the parameters *c* and *m*

Table 6.1. Results of statistical testing - Synthetic data

model	Training sets												Testing sets											
c	3			5			7			9			3			5			7			9		
1 vs. 2	+	+	+	+	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+	-	+	+
1 vs. 3	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+	+	+	+	+	+
1 vs. 4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+
1 vs. 5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

As Figure 6.4 visualizes, in all the models, the increase in the number of clusters leads to the decrease of the training and testing error. Obviously, by having more clusters, the model is able to produce more details of the data leading to the reduction of the training and testing errors. The performance of the proposed models (models 2-5) is better than the one of the original model (model 1). Moreover, the *t*-test run at the confidence level of 0.05 shows the improvement of the proposed models vis-à-vis the original model, see Table 6.1. The first columns concern m = 1.1, gray columns refer to m = 2.0, and the darker columns deal with m = 2.8. The plus sign denotes that the results produced by the two models are statistically different (the differences are statistically significant). In general, the performance of the fuzzy models endowed with the interaction layer is significantly improved in the most cases, and the best results are produced by model 4 and model 5. In other words, we can use a lower number of clusters (rules) to obtain the same performance as in the case of some models composed of the larger number of rules.

Now we present the results for the granular augmentation of the fuzzy models as discussed. In particular, we show the plots of the coverage versus the specificity and the corresponding AUC (V) values. As can be seen from Figure 6.5 and Figure 6.6, the AUC (V) values corresponding to the coverage versus the specificity of the proposed models (models 2-5) are higher than those reported for the original models (model 1).

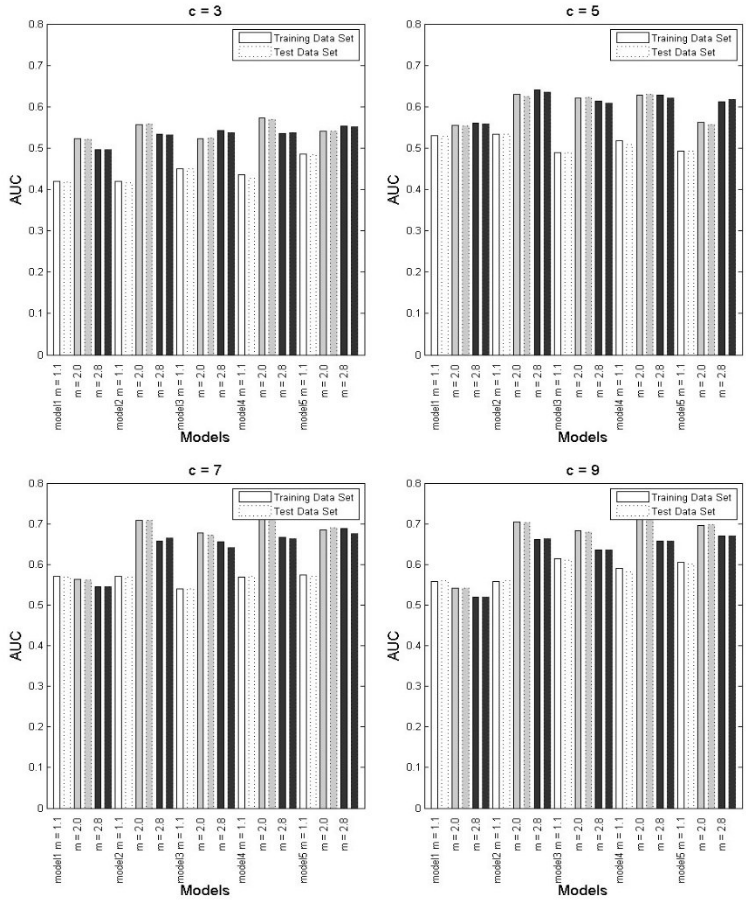


Figure 6.5. AUC corresponding to the coverage versus the specificity in granular scenario (a)

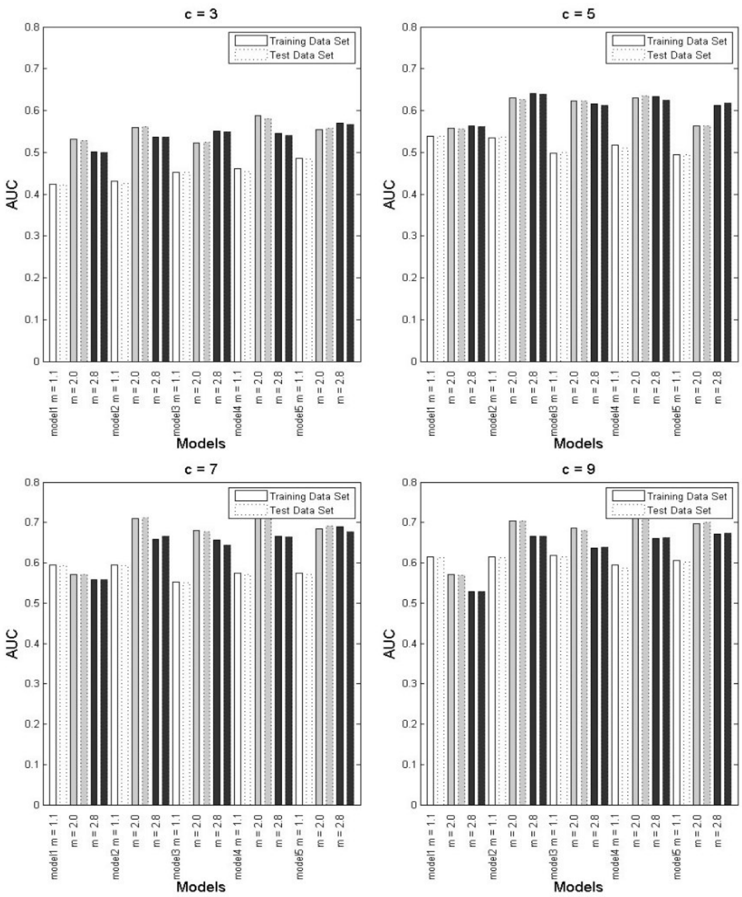


Figure 6.6. AUC corresponding to the coverage versus the specificity in granular scenario (b)

6.2 Granular Interval Prototypes and Parameters of Fuzzy Models[e]

Considering the uncertainty and inaccuracy of prototypes and parameters, we make these numeric values embraced by information granules—that is, we make the fuzzy model granular—in the sense that its original numeric parameters and constructions (prototypes) are generalized to become granular parameters and prototypes. These parameters and prototypes are made granular in the sequel resulting in a granular fuzzy model. With the use of information granulation of the parameters and prototypes of fuzzy models, the resulting granular fuzzy model offers higher level

[e] A version of this chapter has been accepted to published as [132].

tolerance to noise and modeling errors and helps produce results of practical relevance. Therefore, Granular model augments the existing numeric results by generating prediction intervals so that one can expect where the real-world outcome is going to be located. The practical relevance of non-numeric prediction has been recognized and emphasized in system modeling in the past; one can refer here to the use of prediction intervals in power systems reported in [119][120] or allude to interval-like prediction in models of linear regression. Information granularity along with its level of granularity is regarded as a certain design asset facilitating the evaluation of the fuzzy model.

In the study, we propose a two-phase development process of granular fuzzy models, which, in our opinion, is both legitimate and sound from the perspective of the introduced optimization criteria as well as the underlying optimization. The first phase concerns the development of a fuzzy model. Here we resort ourselves to the spectrum of existing well-established design practices of fuzzy models. No change is being made to the design. We take full advantage of what has been fully documented in the literature and successfully used so far. The second design phase is the crux of the overall construct: here we augmented the already constructed fuzzy model by making its parameters and prototypes granular (represented by information granules). The optimization is guided by the aggregate criterion of coverage and specificity (which is fully in line with a way in which the quality of information granules is quantified vis-à-vis the available numeric evidence).

There are several facets of originality of this study. By bringing the concept, performance measures, and ensuing algorithms of the granular evaluation, we embark on a new and uncharted territory of building and expressing performance of granular fuzzy models in a holistic way. The way of an effective building granular fuzzy models realized on a basis of the existing model [52][121] and an assessment of its performance brings another aspect of originality.

In spite of the presence of fuzzy sets used in the development of fuzzy models, these models manifest as numeric constructs (viz. the output of the fuzzy model is numeric). The radical design departure leading to an evident enhancement of the ensuing model is to admit granular parameters of the model. The original numeric values of the parameter of the fuzzy model (say, prototypes of the clusters $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_c, w_1, w_2, \ldots, w_c$ and parameters of the local models $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_c$) are generalized to the form of information granules $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_c, W_1, W_2, \ldots, W_c, \boldsymbol{A}_1, \boldsymbol{A}_2, .., \boldsymbol{A}_c$ so that they are distributed around the original numeric values . To form granular parameters, we admit a certain level of information granularity (ε) (the details will be discussed in the consecutive sections) and allocate it across the parameters of the model so that a certain optimization criterion becomes optimized.

Such information granules are formed around the original numeric values of the parameters. Alluding to the generic structure of the rule-based model governed by (2.6), instead of numeric values of the parameters we consider information granules located around the numeric values of the parameters. Formally, as a follow-up of the original formulas, we describe the resulting granular model as follows

$$\text{if } \boldsymbol{x} \text{ is } G(B_i(\boldsymbol{x})) \text{ then } \hat{Y}_i = W_i \oplus \boldsymbol{A}_i^T \otimes (\boldsymbol{x} - \boldsymbol{V}_i) \tag{6.12}$$

the capital letters emphasize that the corresponding components are information granules and $G(B_i(\boldsymbol{x}))$ denotes an interval of levels of activation (firing) of the rule generated on a basis of the granular prototypes, viz. $G(B_i(\boldsymbol{x})) = [B_i^-(\boldsymbol{x}), B_i^+(\boldsymbol{x})]$. The operations $\oplus$ and $\otimes$ are carried out for information granules as the calculations in interval arithmetic. Next the output of the model follows a generalizes (6.12) and comes in the form

$$\hat{Y} = \sum_{\substack{i=1 \\ \oplus}}^{c} G(B_i(\boldsymbol{x})) \otimes (W_i \oplus \boldsymbol{A}_i^T \otimes (\boldsymbol{x} - \boldsymbol{V}_i)) \tag{6.13}$$

To admit granular parameters and prototypes, and optimize an allocation of information granularity, two fundamental issues have to be studied in depth:

(i) *a way of characterizing the quality of the granular models* This problem is more advanced that the one encountered in (numeric) fuzzy models. In any possible evaluation of the granular model one has to take this into consideration a fact that a numeric data is confronted with an information granule produced by the granular model. This calls for a prudent quantification of the result.

(ii) *optimization of allocation of information granularity across the fuzzy models* How the level of information granularity (being treated as an essential design asset) can be distributed across the parameters and prototypes of the fuzzy rule-based model to yield the best performance of the granular model is the crux of the design problem.

We note that way of proceeding with a granular fuzzy model exhibits a significant level of generality. We have not committed to any particular formalism of realization of information granules. In the ensuing detailed investigations, we confine ourselves to intervals as this helps focus on the essence of the approach and avoid venturing into computational details.

There are several essential components contributing to the entire construction:

- a level of information granularity ε is provided in advance. It can be regarded as a supplied design asset so that the model is made granular. The values of ε are confined to the unit interval. The higher the values of ε, the more design flexibility is being offered. In the limit when $\varepsilon = 0$ the granular model becomes the original one (numeric model),
- a way of allocation information granularity, viz. a method of making the original numeric parameters granular,
- a way of expressing the quality of the produced granular model.

6.2.1 Granulation of the Parameters of Fuzzy Models

Given the numeric value of the parameter of the local model, say a_{ij}, $i = 1, 2, \ldots, c$, $j = 1, 2, \ldots, n$, we make it granular by admitting a certain interval $[a_{ij}^-, a_{ij}^+]$ formed around the original numeric value. Obviously, because of the granular form of the parameters $A = [a^-, a^+]$, the numeric outputs become effectively granular (more specifically, intervals) $Y = [y^-, y^+]$ whose bounds are described in the following form,

$$\left[y_k^-, y_k^+\right] = \sum_{i=1}^{c} A_i(\boldsymbol{x}_k) \otimes \left[\omega_i \oplus \left[\boldsymbol{a}_i^{-T}, \boldsymbol{a}_i^{+T}\right] \otimes (\boldsymbol{x}_k - \boldsymbol{v}_i)\right] \tag{6.14}$$

Allocation of granularity of parameters

(I) *Uniform and symmetric allocation of granularity*. The first scenario realizes a symmetric and uniform allocation of the level of information granularity by building the bounds of the intervals as follows

$$a_{ij}^- = \begin{cases} \min\left(a_{ij}(1-\varepsilon_0), a_{ij}(1+\varepsilon_0)\right) & \text{if } a_{ij} \neq 0 \\ -\varepsilon_0 & \text{if } a_{ij} = 0 \end{cases} \tag{6.15}$$

$$a_{ij}^+ = \begin{cases} \max\left(a_{ij}(1-\varepsilon_0), a_{ij}(1+\varepsilon_0)\right) & \text{if } a_{ij} \neq 0 \\ \varepsilon_0 & \text{if } a_{ij} = 0 \end{cases} \tag{6.16}$$

where $\varepsilon_0 = \varepsilon/2cn$.

(II) *Nonuniform and asymmetric allocation of granularity*. Here we consider the length of each interval around the numeric values of the parameters is not equal.

$$a_{ij}^- = \begin{cases} \min\left(a_{ij}(1-\varepsilon_{ij}^-), a_{ij}(1+\varepsilon_{ij}^+)\right) & \text{if } a_{ij} \neq 0 \\ \varepsilon_{ij}^- & \text{if } a_{ij} = 0 \end{cases} \tag{6.17}$$

$$a_{ij}^+ = \begin{cases} \max\left(a_{ij}(1-\varepsilon_{ij}^-), a_{ij}(1+\varepsilon_{ij}^+)\right) & \text{if } a_{ij} \neq 0 \\ \varepsilon_{ij}^+ & \text{if } a_{ij} = 0 \end{cases} \tag{6.18}$$

The variables ε_{ij}^- and ε_{ij}^+ are satisfied the following conditions,

$$0 \leq \varepsilon_{ij}^-, \varepsilon_{ij}^+ \leq 1 \tag{6.19}$$

$$\sum_{i=1}^{c}\left(\sum_{j=1}^{n}\varepsilon_{ij}^- + \sum_{j=1}^{n}\varepsilon_{ij}^+\right) = \varepsilon \tag{6.20}$$

6.2.2 Granulation of the Prototypes of Fuzzy Models

The prototype produced by the FCM algorithm is numeric. To generalize it to a granular construct, we build a hyper-rectangle prototype $\boldsymbol{V}_i$. A two-dimensional ($n = 2$)

illustration is displayed in Figure 6.7. In general, we admit the sides of the hypercube to be of different lengths and distributed asymmetrically around the prototypes. The resulting granular prototype is fully described by 2^n parameters (4 in the two-dimensional case).

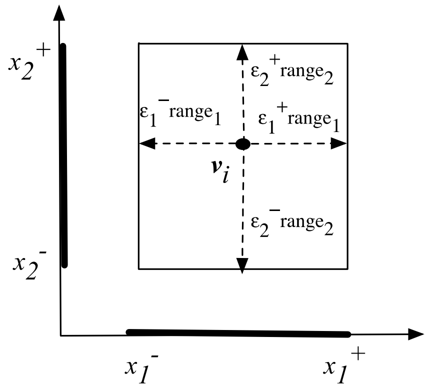


Figure 6.7. Constructing a granular prototype; a two-dimensional example. $range_1$ and $range_2$ are the ranges of values assumed by the corresponding variables x_1 and x_2.

Given a certain data $\boldsymbol{x}$, a level of activation of the i-th rule comes in the form of an interval. Alluding to the way in which this is done in fuzzy models, here we have to express on how to determine a distance between a numeric vector and the hyper rectangular information granule $\boldsymbol{V}_i$. Intuitively, the result should be non-numeric and we should take into consideration a range of possible values assumed by the distance involving their extreme values. An intuitively sound option would be to take the most distant and the closest vertices of $\boldsymbol{V}_i$. This, however, becomes computationally questionable when n attains higher values as we have to consider 2^n vertices of the hyper rectangle to determine the bounds of these distance values. An alternative is to project the hyper rectangle on the corresponding axes and determine the distance between the projected $\boldsymbol{x}$, say $\boldsymbol{x}_j$ and the most distant and the closest point of the interval resulting from the projection of the hyper rectangle on the same j-th input variable, for details refer to Figure 6.7.

We arrive at the detailed formulas

$$A_i^1(\boldsymbol{x}_k) = \begin{cases} \dfrac{1}{\sum_{j=1}^{c} \left(\dfrac{\|\boldsymbol{x}_k - \boldsymbol{v}_i^+\|}{\|\boldsymbol{x}_k - \boldsymbol{v}_j^+\|} \right)^{2/(m-1)}} & \text{if } \boldsymbol{x}_k \notin \boldsymbol{V}_i \\ 1 & \text{if } \boldsymbol{x}_k \in \boldsymbol{V}_i \end{cases} \tag{6.21}$$

$$A_i^2(\boldsymbol{x}_k) = \begin{cases} \dfrac{1}{\sum_{j=1}^{c} \left(\dfrac{\|\boldsymbol{x}_k - \boldsymbol{v}_i^-\|}{\|\boldsymbol{x}_k - \boldsymbol{v}_j^-\|} \right)^{2/(m-1)}} & \text{if } \boldsymbol{x}_k \notin \boldsymbol{V}_i \\ 1 & \text{if } \boldsymbol{x}_k \in \boldsymbol{V}_i \end{cases} \tag{6.22}$$

$$\|\boldsymbol{x}_k - \boldsymbol{v}_i^-\| = \sqrt{\sum_{j=1}^{n} \frac{(x_{kj} - b_{ij}^-)^2}{\sigma_j^2}} \tag{6.23}$$

$$\|\boldsymbol{x}_k - \boldsymbol{v}_i^+\| = \sqrt{\sum_{j=1}^{n} \frac{(x_{kj} - b_{ij}^+)^2}{\sigma_j^2}} \tag{6.24}$$

where σ_j is a standard deviation of the j-th variable while b_{ij}^- and b_{ij}^+ are defined as follows,

$$b_{ij}^- = \begin{cases} x_j^- & \text{if } x_{kj} < x_j^- \\ x_{kj} & \text{if } x_j^- \le x_{kj} \le x_j^+ \\ x_j^+ & \text{if } x_{kj} > x_j^+ \end{cases} \tag{6.25}$$

$$b_{ij}^+ = \begin{cases} x_j^+ & \text{if } x_{kj} < x_j^- \\ x_{kj} & \text{if } x_j^- \le x_{kj} \le x_j^+ \\ x_j^- & \text{if } x_{kj} > x_j^+ \end{cases} \tag{6.26}$$

$$\boldsymbol{v}_i^- = [b_{i1}^-, b_{i2}^-, ..., b_{in}^-] \tag{6.27}$$

$$\boldsymbol{v}_i^+ = [b_{i1}^+, b_{i2}^+, ..., b_{in}^+] \tag{6.28}$$

Then

$$A_i^-(\boldsymbol{x}_k) = \min\left(A_i^1(\boldsymbol{x}_k), A_i^2(\boldsymbol{x}_k)\right) \tag{6.29}$$

$$A_i^+(\boldsymbol{x}_k) = \max\left(A_i^1(\boldsymbol{x}_k), A_i^2(\boldsymbol{x}_k)\right) \tag{6.30}$$

The output interval reads as follows

$$\left[y_k^-, y_k^+\right] = \sum_{i=1}^{c} \left[A_i^-(\boldsymbol{x}_k), A_i^+(\boldsymbol{x}_k)\right] \otimes \left[\left[w_i^-, w_i^+\right] \oplus \boldsymbol{a}_i^T \otimes \left[\boldsymbol{x}_k - \boldsymbol{v}_i^-, \boldsymbol{x}_k - \boldsymbol{v}_i^+\right]\right] \tag{6.31}$$

Allocation of granularity of prototypes

(I) *Uniform and symmetric allocation of granularity*. The first scenario realizes a symmetric and uniform allocation of the level of information granularity by building the bounds of the intervals as follows

$$\boldsymbol{v}_i \rightarrow \boldsymbol{V}_i\left(\left[\varepsilon_0 rang_i^x, \varepsilon_0 rang_i^x, ..., \varepsilon_0 rang_i^x\right]_{1\times n}\right) \tag{6.32}$$

where $range_i^x = x_{\max} - x_{\min}$.

$$\omega_i^- = \min\left(\omega_i - \varepsilon_0 range_i^y, \omega_i + \varepsilon_0 range_i^y\right) \tag{6.33}$$

$$\omega_i^+ = \max\left(\omega_i - \varepsilon_0 range_i^y, \omega_i + \varepsilon_0 range_i^y\right) \tag{6.34}$$

where $\varepsilon_0 = \varepsilon/2c(n+1)$ and $range_i^y = y_{\max} - y_{\min}$.

(II) *Nonuniform and asymmetric allocation of granularity*. Here we consider the length of each interval around the numeric values of the parameters is not equal.

$$\boldsymbol{v}_i \rightarrow \boldsymbol{V}_i\left(\varepsilon_{i1}^{v-} rang_i^x, \varepsilon_{i1}^{v+} rang_i^x, \varepsilon_{i2}^{v-} rang_i^x, \varepsilon_{i2}^{v+} rang_i^x, ..., \varepsilon_{in}^{v-} rang_i^x, \varepsilon_{in}^{v+} rang_i^x\right) \tag{6.35}$$

$$w_i^- = \min\left(w_i - \varepsilon_i^{w-} range_i^y, w_i + \varepsilon_i^{w+} range_i^y\right) \tag{6.36}$$

$$w_i^+ = \max\left(w_i - \varepsilon_i^{w-} range_i^y, w_i + \varepsilon_i^{w+} range_i^y\right) \tag{6.37}$$

The variables ε_{ij}^{v-}, ε_{ij}^{v+}, ε_i^{w-}, and ε_i^{w+} are satisfied the following conditions,

$$0 \le \varepsilon_{ij}^{v-}, \varepsilon_{ij}^{v+}, \varepsilon_{i}^{w-}, \varepsilon_{i}^{w+} \le 1 \tag{6.38}$$

$$\sum_{i=1}^{c}\left(\sum_{j=1}^{n}\varepsilon_{ij}^{v-} + \sum_{j=1}^{n}\varepsilon_{ij}^{v+} + \varepsilon_{i}^{w-} + \varepsilon_{i}^{w+}\right) = \varepsilon \tag{6.39}$$

6.2.3 Granulation of the Parameters and Prototypes of Fuzzy models

Based on the above two strategies, we consider combining the granulation of both parameters and prototypes. The components of granulation, for instance, the granulating parameters, distance and partitions calculation, are same as previous description. In contrast, the calculation of granular output is changed as following expression, which contains all granular components of the rule-based fuzzy model.

$$\left[y_k^-, y_k^+\right] = \sum_{i=1}^{c}\left[A_i^-(\boldsymbol{x}_k), A_i^+(\boldsymbol{x}_k)\right] \otimes \left[\left[w_i^-, w_i^+\right] \oplus \left[\boldsymbol{a}_i^{-T}, \boldsymbol{a}_i^{+T}\right] \otimes \left[\boldsymbol{x}_k - \boldsymbol{v}_i^-, \boldsymbol{x}_k - \boldsymbol{v}_i^+\right]\right] \tag{6.40}$$

In the sequel, we carefully look at the ways of allocating information granularity and the quality of the granular model.

Allocation of granularity of both parameters and prototypes of fuzzy models

(I) *Uniform and symmetric allocation of granularity*.

$$a_{ij}^- = \begin{cases} \min\left(a_{ij}(1-\varepsilon_0), a_{ij}(1+\varepsilon_0)\right) & \text{if } a_{ij} \ne 0 \\ -\varepsilon_0 & \text{if } a_{ij} = 0 \end{cases} \tag{6.41}$$

$$a_{ij}^+ = \begin{cases} \max\left(a_{ij}(1-\varepsilon_0), a_{ij}(1+\varepsilon_0)\right) & \text{if } a_{ij} \ne 0 \\ \varepsilon_0 & \text{if } a_{ij} = 0 \end{cases} \tag{6.42}$$

$$\boldsymbol{v}_i \to \boldsymbol{V}_i\left(\left[\varepsilon_0 rang_i^x, \varepsilon_0 rang_i^x, \ldots, \varepsilon_0 rang_i^x\right]_{1\times n}\right) \tag{6.43}$$

where $\varepsilon_0 = \varepsilon / 2(cn + c(n+1))$.

(II) *Nonuniform and asymmetric allocation of granularity*. Here we consider the length of each interval around the numeric values of the parameters is not equal.

$$a_{ij}^{-}=\begin{cases}\min\left(a_{ij}(1-\varepsilon_{ij}^{a-}),a_{ij}(1+\varepsilon_{ij}^{a+})\right) & \text{if } a_{ij}\neq 0\\ \varepsilon_{ij}^{a-} & \text{if } a_{ij}=0\end{cases} \tag{6.44}$$

$$a_{ij}^{+}=\begin{cases}\max\left(a_{ij}(1-\varepsilon_{ij}^{a-}),a_{ij}(1+\varepsilon_{ij}^{a+})\right) & \text{if } a_{ij}\neq 0\\ \varepsilon_{ij}^{a+} & \text{if } a_{ij}=0\end{cases} \tag{6.45}$$

$$\boldsymbol{v}_i \to \boldsymbol{V}_i\left(\varepsilon_{i1}^{v-}rang_i^x,\varepsilon_{i1}^{v+}rang_i^x,\varepsilon_{i2}^{v-}rang_i^x,\varepsilon_{i2}^{v+}rang_i^x,\ldots,\varepsilon_{in}^{v-}rang_i^x,\varepsilon_{in}^{v+}rang_i^x\right) \tag{6.46}$$

$$w_i^{-}=\min\left(w_i-\varepsilon_i^{w-}range_i^y,w_i+\varepsilon_i^{w+}range_i^y\right) \tag{6.47}$$

$$w_i^{+}=\max\left(w_i-\varepsilon_i^{w-}range_i^y,w_i+\varepsilon_i^{w+}range_i^y\right) \tag{6.48}$$

The variables $\varepsilon_i^{w-},\varepsilon_i^{w+}$, ε_{ij}^{a-}, ε_{ij}^{v-}, ε_{ij}^{a+} and ε_{ij}^{v+} are satisfied the following conditions,

$$0\le \varepsilon_{ij}^{a-},\varepsilon_{ij}^{a+},\varepsilon_{ij}^{v-},\varepsilon_{ij}^{v+},\varepsilon_i^{w-},\varepsilon_i^{w+}\le 1 \tag{6.49}$$

$$\sum_{i=1}^{c}\left(\sum_{j=1}^{n}\varepsilon_{ij}^{a-}+\sum_{j=1}^{n}\varepsilon_{ij}^{a+}+\sum_{j=1}^{n}\varepsilon_{ij}^{v-}+\sum_{j=1}^{n}\varepsilon_{ij}^{v+}+\varepsilon_i^{w-}+\varepsilon_i^{w+}\right)=\varepsilon \tag{6.50}$$

In (II) case given the large number of values to be determined, it can consider resorting to some evolutionary optimization techniques, such as particle swarm optimization (PSO), genetic algorithm (GA), and differential evolution (DE).

6.2.4 Experimental Studies

A two-variable nonlinear function comes in (6.11), where x_1 and x_2 are two independent variables assuming values within the range [1, 5]. The data set is composed of 900 (30×30) input-output data pairs where each input is distributed according to the uniform random distribution in [1, 5]. The data set is split into a training set (70%) and testing set (30%). The initial values of ε are used in the range [0, 1] with a step 0.1.

To quantify the performance of the granular fuzzy rule-based models, we implement coverage, specificity and the global index as described in (2.29), (2.32), and (2.39), respectively.

Figure 6.8 displays the relationships between coverage and specificity obtained for different granular models when allocating information granularity. The lines with circles represent a uniform allocation of information granularity while the lines with pentagram correspond to the case where information granularity has been optimized. As expected, considering the conflicting nature of the coverage and specificity criteria, the increase in the specificity comes at the expense of the decreasing values of coverage. Table 6.2 includes V values corresponding to the two scenarios of allocation of information granularity. Some general tendencies are apparent. The V values of the granular models with the optimized granular parameters are higher with those produced when realizing a uniform allocation of information granularity. Furthermore, one can quantify the improvement delivered by the PSO method. The obtained curves display different shapes and in some cases. For example, Figure 4 for $c = 5$ and $c = 9$, there are relatively flat regions of the curve meaning that one can increase coverage (by increasing the value of ε) not sacrificing much the specificity of the results. With the optimized model, some of the coverage and specificity values have improved meaning that the model leads to the better coverage with the similar values of specificity. or in some cases the coverage is not enhanced too much but the specificity has some improvement. Moreover, in some cases (for instance, when $c = 2$), after optimization, the specificity has shown a significant improvement while the coverage is reduced, and the points in the coverage-specificity coordinates are positioned very closely. A possible reason is that in those models, the allocation of information granularity has a strong impact on the specificity improvement and a far less visible effect on coverage enhancement. As a result, one improved specificity with sacrificing the coverage.

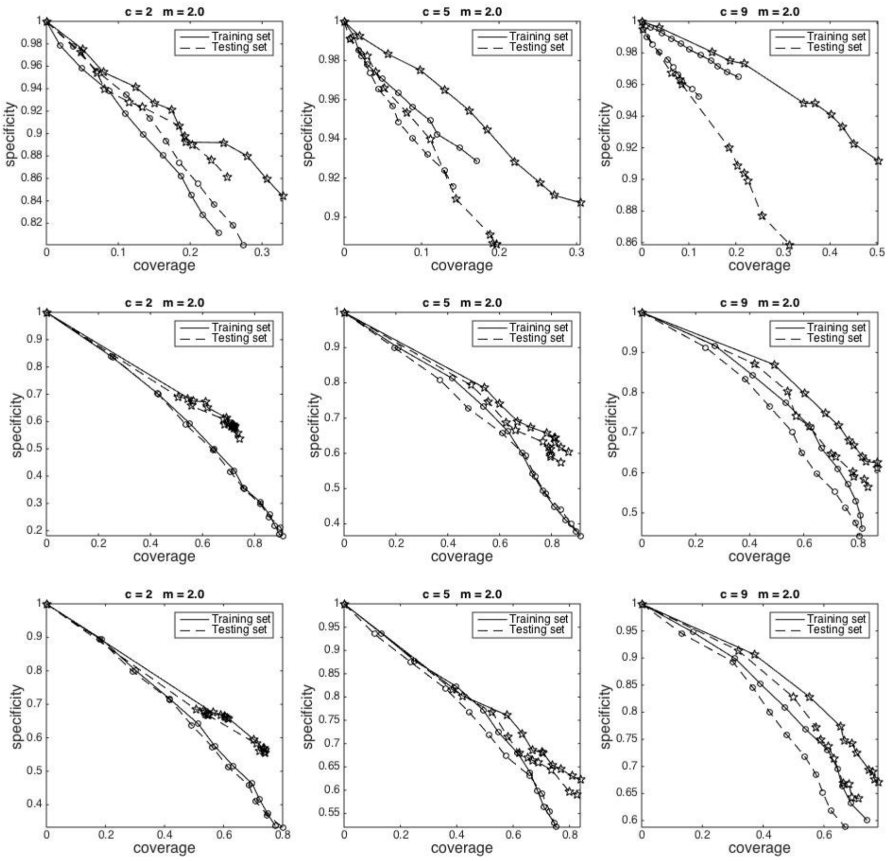


Figure 6.8.Coverage versus the specificity obtained by three ways of allocation of information granularity across the parameters of the model: granulation of parameters of linear models in the condition part – first row, granulation of prototypes – second row, granulation of parameters and prototypes – third row.

Table 6.2. Comparison of *V* values of granular models after optimization of allocation of information granularity

Granularity	$c = 2$		$c = 5$		$c = 9$	
	Training	Testing	Training	Testing	Training	Testing
Parameters	0.1721	0.1361	0.1587	0.0953	0.2904	0.1476
Prototypes	0.4120	0.3981	0.4927	0.4587	0.5104	0.4553
Para. & Prot.	0.4058	0.3809	0.4661	0.4355	0.4905	0.4328

6.3 Granular Fuzzy Sets the Conclusion Part Parameters of Fuzzy Models[f]

As mentioned in the introduction, information granules augmenting fuzzy models can be formalized in different ways by invoking intervals, fuzzy sets, rough sets, shadowed sets, and probabilistic sets—to name a few alternatives. We have considered several development techniques for granular models that numeric models are granulated yielding granular intervals. In this chapter, to illustrate another promising alternative, we consider formalizing information granules in a formation of fuzzy sets for TS fuzzy rule-based models. We note that a granular model augmented by fuzzy sets offer visible advantages over the augmentation involving interval information granules. Moreover, fuzzy sets help quantify linguistic variables, which is helpful in improving the interpretability of granular fuzzy models. In light of this observation, it becomes apparent that this form of information granularity delivers an enhancement of the original fuzzy model.

The underlying objective of this study is to establish a concept of granular fuzzy models, highlight their key features and advantages as well as develop a comprehensive and efficient design strategy of such models. The essence of the proposed design is that the granular model is being built on a basis of the already existing fuzzy model so that the resulting construct is built in an efficient way and may fully benefit from the already constructed fuzzy model. The original performance index (involving two important criteria pertinent to information granules, namely coverage and specificity) is prudently formed to navigate the construction of the granular model.

In this study, we concentrate on TS fuzzy rule-based model as a framework to develop granular fuzzy models. The objective and originality of this study is to establish an appropriate (optimal) way of allocation of information granularity around the parameters of fuzzy models, so that the parameters of generic fuzzy models can be

[f] A version of this chapter has been published as [133].

augmented to the fuzzy sets form. In the sequel, the output of the granular model becomes a fuzzy set. Two methods are implemented to evaluate the performance of the established granular fuzzy model. Firstly, we involve the principle of justifiable information granularity that assess the coverage and specificity of the resulting fuzzy set.

6.3.1 Fuzzy Rule-Based System Modeling

Here we consider granulating the parameters in the conclusion parts of fuzzy rules, therefore, the function $f_i(x)$ in the conclusion parts is slightly different from above two sections. The series of "if-then" statements (rules) assuming the following form and modeling a relationship between input variables $\boldsymbol{x}$ and output variable

$$\text{If } \boldsymbol{x} \text{ is } A_i \text{ then } y_i \text{ is } f_i(\boldsymbol{x}) = p_{i0}+p_{i1}x_1+\ldots+p_{in}x_n \tag{6.51}$$

where y_i is a linear function of the input variables parameters p_{i0}, p_{i1}, …, p_{in}, and n is the number of input variables. The following process of computing the output of the model are the same as described above in (2.8).

6.3.2 Formation of Granular Fuzzy Model

We augment the generic topology of the fuzzy model to form a granular fuzzy model, in which the parameters p_{i0}, p_{i1}, …, p_{ij}, …, p_{in} of the linear functions are extended to some fuzzy sets (for instance, triangular fuzzy sets and parabolic fuzzy sets). The parametric version of such fuzzy numbers can be expressed as follows

$$P_{ij} = [p_{ij}^-, p_{ij}, p_{ij}^+] \tag{6.52}$$

where p_{ij}^-, p_{ij}^+ (j = 0, 1, …, n) are the lower and upper bounds of the corresponding fuzzy sets, which is calculated as follows,

$$p_{ij}^- = \begin{cases} \min\left(p_{ij}\left(1-\varepsilon_{ij}^-\right), p_{ij}\left(1+\varepsilon_{ij}^+\right)\right) & p_{ij} \neq 0 \\ -\varepsilon_{ij}^- & p_{ij} = 0 \end{cases} \tag{6.53}$$

$$p_{ij}^{+} = \begin{cases} \max\left(p_{ij}\left(1-\varepsilon_{ij}^{-}\right), p_{ij}\left(1+\varepsilon_{ij}^{+}\right)\right) & p_{ij} \neq 0 \\ \varepsilon_{ij}^{+} & p_{ij} = 0 \end{cases} \quad (6.54)$$

where ε_{ij}^{-}, ε_{ij}^{+} are levels (allocation) of information granularity allocated to the corresponding numeric parameters.

As a result, the granular rules are augmented as following form,

$$\text{If } \boldsymbol{x} \text{ is } A_i \text{ then } Y_i = P_{i0} \oplus P_{i1} \otimes x_1 \oplus \ldots \oplus P_{in} \otimes x_n \quad (6.55)$$

where Y_i is a fuzzy set output of the i-th rule, and the operation $\oplus$ and $\otimes$ denote addition and multiplication of fuzzy numbers.

Following the rules of fuzzy arithmetic, the resulting fuzzy set Y_{ik} associated with the i-th rule for input $\boldsymbol{x}_k$ is expressed as a fuzzy number with the following membership function,

$$Y_{ik} = \left[\; lb_{ik},\, p_{i0} + \sum_{j=1}^{n} p_{ij} x_{jk},\, ub_{ik} \;\right] \quad (6.56)$$

where the lower and upper bounds of the membership function are expressed as

$$lb_{ik} = \min\left(p_{i0}^{-},\; p_{i0}^{+}\right) + \sum_{j=1}^{n} \min\left(p_{ij}^{-} x_{jk},\; p_{ij}^{+} x_{jk}\right) \quad (6.57)$$

$$ub_{ik} = \max\left(p_{i0}^{-},\; p_{i0}^{+}\right) + \sum_{j=1}^{n} \max\left(p_{ij}^{-} x_{jk},\; p_{ij}^{+} x_{jk}\right) \quad (6.58)$$

The minimum and maximum operators are applied individually to each coordinate (variable), $j = 1,2\ldots n$. Then, the original formula is described as the resulting fuzzy set of output in the form

$$\left[\, l_k, \hat{y}_k, u_k \,\right] = \sum_{i=1}^{c} A_{ik} \otimes Y_{ik} \quad (6.59)$$

where

$$l_k = \sum_{i=1}^{c} A_{ik} lb_{ik} \tag{6.60}$$

$$\hat{y}_k = \sum_{i=1}^{c} A_{ik} \left(p_{i0} + \sum_{j=1}^{n} p_{ij} x_{jk} \right) \tag{6.61}$$

$$u_k = \sum_{i=1}^{c} A_{ik} ub_{ik} \tag{6.62}$$

To restrict the fuzzy set to positioned within the output space, the following clipping operation is introduced,

$$l_k^* = \begin{cases} l_k & y_{\min} \le l_k \le y_{\max} \\ y_{\min} & l_k < y_{\min} \\ y_{\max} & l_k > y_{\max} \end{cases} \tag{6.63}$$

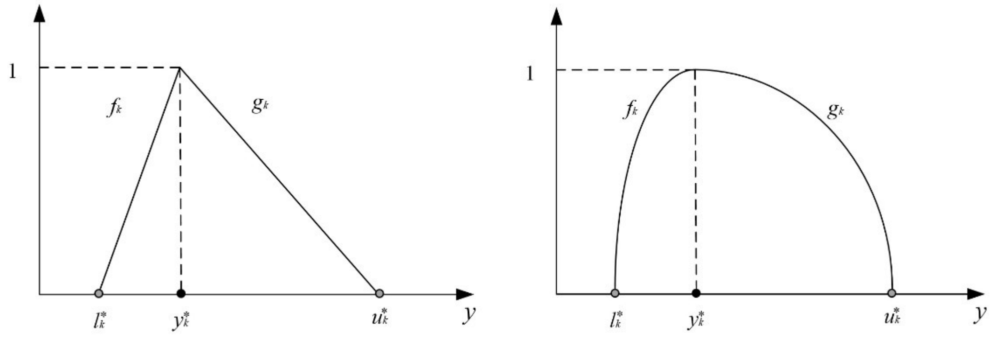
$$u_k^* = \begin{cases} u_k & y_{\min} \le u_k \le y_{\max} \\ y_{\min} & u_k < y_{\min} \\ y_{\max} & u_k > y_{\max} \end{cases} \tag{6.64}$$

$$y_k^* = \begin{cases} \hat{y}_k & l_k^* \le \hat{y}_k \le u_k^* \\ l_k^* & \hat{y}_k < l_k^* \\ u_k^* & \hat{y}_k > u_k^* \end{cases} \tag{6.65}$$

where $y_{\max}$ and $y_{\min}$ are the bounds of the output space.

As a result, any granular fuzzy model yields a result coming in a form of a fuzzy set. The parametric membership functions of triangular fuzzy sets and parabolic fuzzy sets are expressed in terms of left- and right- hand parametric bounded functions f_k and g_k, as shown in Figure 6.9. The output fuzzy set is characterized by the following triple

$$\hat{Y}_k = \left[l_k^*, y_k^*, u_k^* \right] \tag{6.66}$$



(a) (b)

Figure 6.9. Examples of membership functions of (a) triangular fuzzy set, (b) parabolic fuzzy set.

As the result of the granular fuzzy model is an information granule (fuzzy set), its performance has to be carefully assessed vis-à-vis the granules in the form of intervals. The coverage measures the membership values of the fuzzy sets as described in (2.31), and the specificity evaluates the size of the fuzzy sets as in (2.33). As we consider the triangular membership function and parabolic membership function here, the resulting specificity computed with the use of (2.33) concerning triangular membership function and parabolic membership function,

$$sp_k = 1 - \frac{1}{2}\frac{u_k^* - l_k^*}{range} \tag{6.67}$$

$$sp_k = 1 - \frac{2}{3}\frac{u_k^* - l_k^*}{range} \tag{6.68}$$

Here we implement the formula (2.37) to combine coverage and specificity criteria, because it is able to not only optimize the two criteria at the same time but also control the impact of specificity. The global performance index is calculated as (2.39).

ε is a predetermined level of information granularity and used to restrict the allocation of information granularity ε_{ij}^- and ε_{ij}^+ within a certain range as following condition expressing an overall balance of information granularity,

$$\sum_{i=1}^{c}\sum_{j=0}^{n}\left(\varepsilon_{ij}^{-}+\varepsilon_{ij}^{+}\right)=c\varepsilon \tag{6.69}$$

6.3.3 Experimental Studies

A two-variable nonlinear function comes in the same as (6.11). The settings of generate data sets are same as in Chapter 6.2. The fuzzification coefficient (m) used in the FCM algorithm assumes three selected values, namely 1.1, 2.0, and 2.9. The predetermined values of ε is set as 0.1, 0.2, …, 0.5. The search space of each element ε_{ij}^{-} and ε_{ij}^{+} is set within the range [0, 1].

The global evaluation of the granular model (expressed in terms of V) is presented in Table 6.3-Table 6.8. To optimize allocation of information granularity, we maximize the values of Q in (2.37) and setting β = 1. Here we present three different ways to allocate the information granularity. First, no optimization is considered - we allocate information granularity symmetrically and at the same level. Then PSO and DE algorithm optimize the allocation process. Contrasting the results of constant (non-optimized) allocation and optimal allocation, there is an improvement in the range of 4%~31%. This indicates that the optimization of allocation plays an important role. Moreover, by comparing the results produced by the PSO and DE, the differences in their performance are negligible positioned in the range -0.02%~0.5%. Some illustration of the performance of PSO and DE reported in successive generations is presented in Figure 6.10. The computing overhead of the two approaches is similar however PSO is slightly faster than the DE in producing the best results. In conclusion, we note that PSO algorithm exhibits a comparatively limited advantage over the DE mechanism.

Table 6.3. Values of V for triangular membership functions and m = 1.1

Optimization	Data	c = 3	5	7	9	11	13	15
None	Training	0.4763	0.5063	0.6431	0.7646	0.7960	0.7933	**0.8575**
	Testing	0.4996	0.5183	0.6482	0.7614	0.7903	0.7875	**0.8466**
PSO	Training	0.6147	0.6490	0.7451	0.8269	0.8451	0.8440	**0.8921**
	Testing	0.6235	0.6420	0.7298	0.8227	0.8370	0.8373	**0.8804**

DE	Training	0.6140	0.6490	0.7451	0.8267	0.8449	0.8438	**0.8918**
	Testing	0.6282	0.6420	0.7298	0.8226	0.8369	0.8372	**0.8801**

Table 6.4. Values of V for triangular membership functions and m = 2.0

Optimization	Data	*c*						
		3	5	7	9	11	13	15
None	Training	0.3889	0.4546	0.4388	0.4743	0.5305	0.5768	**0.5882**
	Testing	0.4100	0.4384	0.4350	0.4545	0.5093	0.5426	**0.5610**
PSO	Training	0.5130	0.5609	0.5595	0.5881	0.6468	0.6671	**0.6923**
	Testing	0.5138	0.5400	0.5468	0.5655	0.6203	0.6315	**0.6582**
DE	Training	0.5130	0.5610	0.5573	0.5876	0.6465	0.6651	**0.6902**
	Testing	0.5138	0.5400	0.5462	0.5672	0.6204	0.6301	**0.6573**

Table 6.5. Values of V for triangular membership functions and m = 2.9

Optimization	Data	*c*						
		3	5	7	9	11	13	15
None	Training	0.4280	0.4044	0.4476	0.5587	0.5834	0.5822	**0.6467**
	Testing	0.4077	0.3912	0.4312	0.5332	0.5619	0.5429	**0.5920**
PSO	Training	0.5381	0.5146	0.5513	0.6334	0.6449	0.6508	**0.7004**
	Testing	0.5141	0.5008	0.5299	0.6101	0.6241	0.6142	**0.6442**
DE	Training	0.5369	0.5147	0.5509	0.6329	0.6431	0.6498	**0.6988**
	Testing	0.5143	0.5006	0.5292	0.6090	0.6231	0.6136	**0.6423**

Table 6.6. Values of V for parabolic membership functions and m = 1.1

Optimization	Data	*c*						
		3	5	7	9	11	13	15
None	Training	0.5800	0.6094	0.7346	0.8424	0.8585	0.8563	**0.9015**
	Testing	0.5947	0.6165	0.7334	0.8378	0.8495	0.8482	**0.8929**
PSO	Training	0.7232	0.7541	0.8300	0.8947	0.9037	0.9045	**0.9415**
	Testing	0.7275	0.7417	0.8103	0.8864	0.8853	0.8913	**0.9276**
DE	Training	0.7200	0.7531	0.8289	0.8938	0.9034	0.9027	**0.9404**
	Testing	0.7274	0.7418	0.8040	0.8854	0.8856	0.8907	**0.9274**

Table 6.7. Values of V for parabolic membership functions and m = 2.0

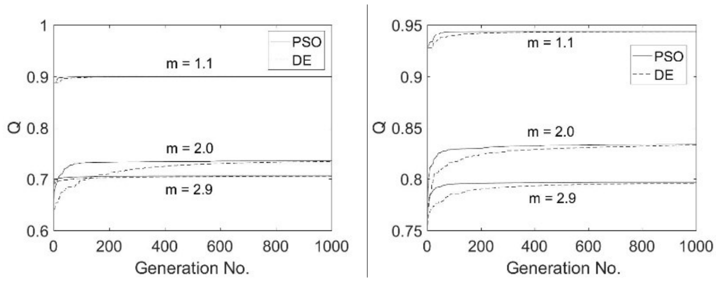
Optimization	Data	*c*						
		3	5	7	9	11	13	15
None	Training	0.4890	0.5596	0.5465	0.5832	0.6386	0.6819	**0.6997**
	Testing	0.5006	0.5363	0.5417	0.5641	0.6133	0.6523	**0.6709**
PSO	Training	0.6151	0.6652	0.6705	0.7016	0.7549	0.7702	**0.7977**
	Testing	0.6020	0.6341	0.6495	0.6779	0.7245	0.7337	**0.7633**

DE	Training	0.6152	0.6650	0.6673	0.6989	0.7542	0.7679	**0.7960**
	Testing	0.6020	0.6365	0.6521	0.6752	0.7242	0.7332	**0.7608**

Table 6.8. Values of *V* for parabolic membership functions and $m = 2.9$

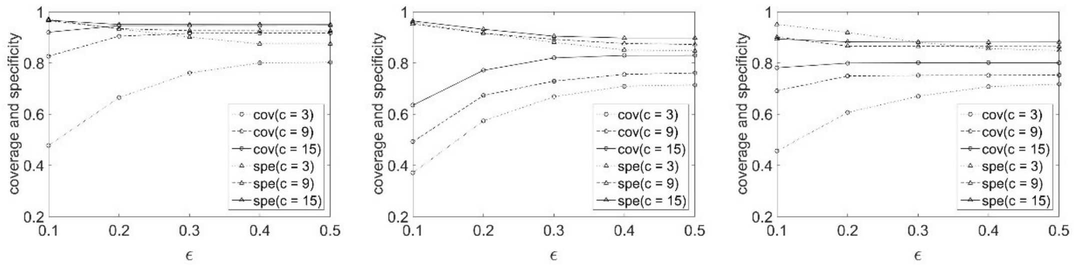
Optimization	Data	*c*						
		3	5	7	9	11	13	15
None	Training	0.5118	0.5033	0.5443	0.6496	0.6639	0.6671	**0.7086**
	Testing	0.4941	0.4868	0.5278	0.6250	0.6443	0.6350	**0.6703**
PSO	Training	0.6276	0.6202	0.6524	0.7347	0.7459	0.7506	**0.7939**
	Testing	0.6026	0.6000	0.6280	0.7069	0.7211	0.7119	**0.7465**
DE	Training	0.6269	0.6201	0.6520	0.7345	0.7440	0.7488	**0.7918**
	Testing	0.6028	0.6006	0.6287	0.7068	0.7192	0.7127	**0.7449**

Furthermore, as shown in the tables, the number of clusters and fuzzification coefficients significantly impact the overall performance. Not surprising, the more number of clusters, the higher values of *V* index are, so having more clusters, the model becomes more capable to capture more details of the data, which happens not only in the numeric fuzzy models but also resulting in granular fuzzy models. In addition, when the number of cluster is relatively lower (in the range 3-7), the performance of the model improves remarkably according to the increasing number of clusters. Whereas, when the number of clusters is higher (greater than 9), the enhancement of performance with more clusters becomes slight. The fuzzification coefficient has notable influence on the membership function A_i used to construct the fuzzy model. Generally, the *V* index is higher when the fuzzification coefficient is small (say 1.1), since error of the fuzzy models with $m = 1.1$ is lower than other two cases. Above all, it seems pertinent to notice that the performance of the granular fuzzy model is considerably depends on the performance of the established fuzzy model. In addition, it is visible that the performance of parabolic membership function model is better than the triangular membership function model. Due to the characteristic of the membership functions, for the same fuzzy numbers, parabolic membership function usually has higher membership grades than triangular fuzzy sets, so the coverage is higher naturally. Despite the specificity of parabolic fuzzy sets is relatively lower than the triangular fuzzy sets, the advantage of coverage still dominates the overall performance.



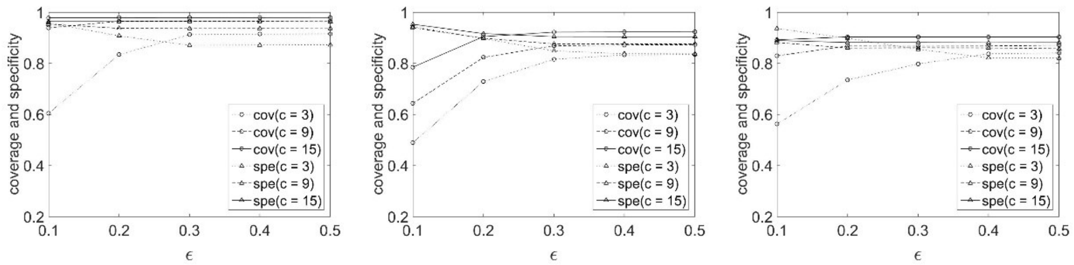
(a) (b)

Figure 6.10. Values of fitness function in successive generations for triangular membership c = 15 and ε = 0.5: (a) triangular membership function model, (b) parabolic membership function model.



(a) (b) (c)

Figure 6.11. Coverage and specificity for triangular membership function model with PSO optimization– Training data: (a) m = 1.1, (b) m = 2.0, (c) m = 2.9.

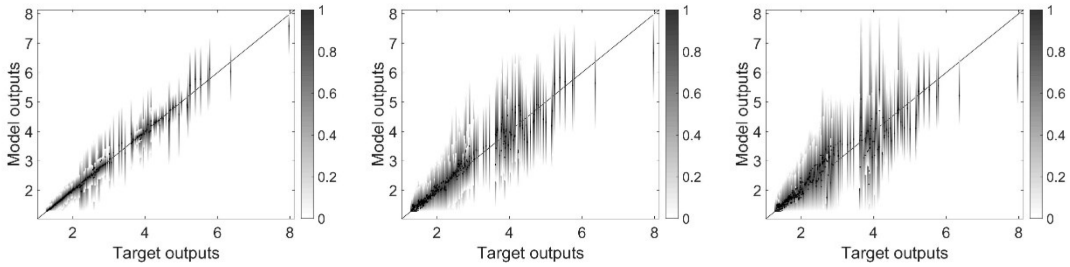


(a) (b) (c)

Figure 6.12. Coverage and specificity criteria for parabolic membership function model with PSO optimization– Training data: (a) m = 1.1, (b) m = 2.0, (c) m = 2.9.

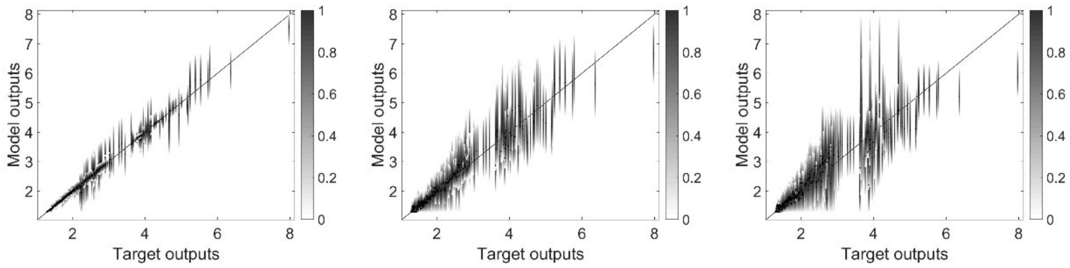
To visualize more results produced by the granular fuzzy model, the coverage and specificity indexes obtained for each model after the completion of optimal allocation are shown in Figure 6.11 and Figure 6.12. Figure 6.13 and Figure 6.14 display some fuzzy sets. The coverage increases while specificity gets lower with the increasing values of the predetermined values of information granularity ε. Furthermore, the number of clusters and the fuzzification coefficient exhibits a direct impact on the coverage index. The coverage performance is better when the number of clusters is higher or fuzzification coefficient is lower, since in these cases, the errors of the established numeric fuzzy models becomes relatively smaller. In contrast, the impact on specificity index due to the number of clusters or fuzzification coefficient is much lower. Therefore, we conclude that predetermined information granularity ε is the most important factor affecting the specificity criterion. In addition, comparing the granular models in case of two types of membership functions (triangular and parabolic), it is worth noting that in the vast majority of the cases, the coverage and specificity implied by parabolic fuzzy sets yields better results. This conclusion can be confirmed by looking at Figure 6.13 and Figure 6.14. In those figures, we use gradient color to show the membership grades. Darker color relates to high value of membership grade. The solid diagonal line represents the model outputs are exactly the same as the target outputs. For the many of fuzzy set outputs come from parabolic membership function models, the color on the diagonal is deeper and the range of the fuzzy sets is shorter, so both the coverage and specificity is better. A possible reason behind this is as follows. Contrasting the membership grades of triangular and parabolic fuzzy sets for the same fuzzy number, the membership grades of parabolic membership function are higher, so the parabolic fuzzy sets are more efficient to provide higher coverage at lower specificity.

The values of ε_{ij} ($\varepsilon_{ij} = \varepsilon_{ij}^{-} + \varepsilon_{ij}^{+}$) allocated across the variables and rules (clusters) are shown as circles in Figure 6.15 and Figure 6.16. The different radii of the circles represent different values of ε_{ij}. The larger the radius, the higher the value of allocated information granularity. Generally, information granularity allocation of triangular membership function model and parabolic membership function model exhibit a similar tendency, and the fuzzification coefficient has significant impact on the allocation. When m = 1.1, the information granularity is allocated on almost every variable, whereas, when m = 2.0 and m = 2.9, the information granularity is mostly allocated to a single variable, and others assume values close to zero.



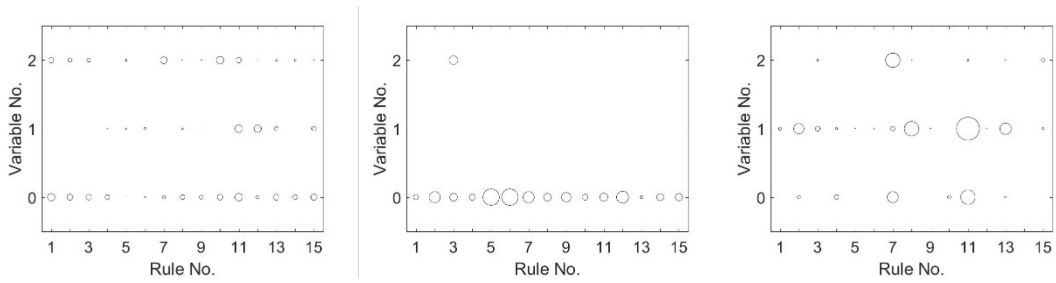
(a) (b) (c)

Figure 6.13. Output y for triangular fuzzy sets c = 15 and ε = 0.5 with PSO optimization– Training data: (a) m = 1.1, (b) m = 2.0, (c) m = 2.9.



(a) (b) (c)

Figure 6.14. Output y for parabolic fuzzy sets c = 15 and ε = 0.5 with PSO optimization– Training data: (a) m = 1.1, (b) m = 2.0, (c) m = 2.9.



(a) (b) (c)

Figure 6.15. Plot of allocated levels of information granularity in case of triangular fuzzy sets c = 15 and ε = 0.5 with PSO optimization: (a) m = 1.1, (b) m = 2.0, (c) m = 2.9.

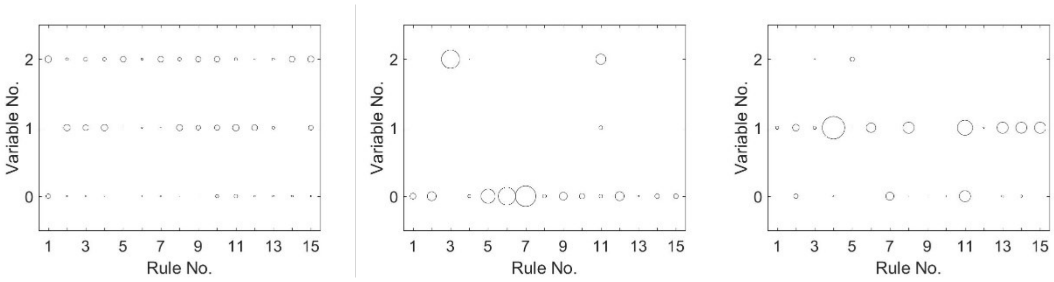


Figure 6.16. Plots of allocated levels of information granularity in case of parabolic fuzzy sets c = 15 and ε = 0.5 with PSO optimization: (a) m = 1.1, (b) m = 2.0, (c) m = 2.9.

6.4 Summary

In this chapter, we developed several granular fuzzy rule-based models by constructing varieties of information granules among the different components of fuzzy rule-based models. In Chapter 6.1, we have augmented granular fuzzy rule-based models by two steps. First, an extended, versatile structure of the rules produced thorough the transformation of condition fuzzy sets is built, which leads to carefully structured membership functions facilitating the overall mapping realized by the rules. The deformation of the original membership functions of the condition part is useful to evaluate a nature and a strength of interaction among the rules, which helps enhance the performance of the model. Second, for the already constructed rules, the model is generalized to its granular version where the transformation matrix is extended to its granular (interval-valued) version. In Chapter 6.2, we augmented numeric fuzzy models to granular by forming information granules around numeric values of the parameters and prototypes of the models. In Chapter 6.3, we have developed the conceptual generalization of fuzzy rule-based models in the form of granular fuzzy sets and presented a comprehensive design by carrying out an optimal allocation of information granularity across the numeric parameters of the conclusion parts of the original model.

The models (and it can be any other models, in general) are evaluated in granular concept brings another more general perspective at the comprehensive evaluation of models and enriches a look at system modeling. We showed that the coverage and specificity measures serve as the two essential measures quantifying a well-rounded way of expressing the quality of the granular model. The characterization of the granular model in terms of its coverage-specificity relationships or the global descriptor coming in the form of the *AUC* measure becomes beneficial to a holistic assessment of the quality of the rule-based models. These two components are crucial in the evaluation of the performance of the granular fuzzy rule-based models. The achieving the tradeoff is facilitated by bringing the parameter granular using which one strikes a sound compromise between the coverage and specificity requirements. The optimization process is guided by the global performance index, which includes the coverage and specificity of the obtained granular results. So that the granular output could cover target evidence as specifically as possible, which makes the granular output of the fuzzy rule-based model of higher practical relevance.

Chapter 7

Granular Output of Fuzzy Rule-Based Models[g]

Granular input space and internal architecture issues are studied in chapters 5 and 6, and information granularity is directly allocated around the numeric input variables and parameters values. This chapter introduces a more straightforward and concise approach that the allocation of information granularity in output space is considered to develop. The output space is typically one-dimensional as the common architectures involve multi-input single-output models. It is nearly useless to simply allocate the information granularity around the only one coordinate of output variable. So, we expect a mechanism that can allocate the information granularity according to the individual output instance, which is, furthermore, a nonlinear mapping from an individual output instance to a granular allocation of output.

The output of numeric model is made granular by associating information granularity to the numeric outputs produced by the original model. The granular output space of models plays a vital role because it has several advantages over other two categories. The allocation of information granularity to input space is quite related to the quality of the model one has started with and the sensitivity of the input variables. Thus, if the numeric model has been constructed, the enhancement of the granular model would be limited. As to the granular the constructions of models, the performance of granular model could be improved by augmentation of parameters and constructive elements with the allocation of granularity, however, in many cases, allocation of information granularity is quite complicated since the parameters and elements are usually represented in multi-dimensions. The curse of dimensionality would likely bring the unacceptable computation consumption of the allocation. In contrast, the granular output space of models is more straightforward and concise in the implementation, because the dimension of the output space is generally low (being

[g] A version of this chapter has been accepted to published as [134].

one-dimensional), and it is also beneficial to remedy some deficiencies of the numeric output.

For the clarity and conciseness of the presentation, we focus on the realization of information granules in the form of intervals, however the underlying conceptual framework is equally suitable to cope with other formal realizations of information granules (say, fuzzy sets, rough sets). It is worth emphasizing that, even though the granular output we produced in this study is in the form of intervals, it is fundamentally different from the proposed interval, possibilistic regression algorithms or its related works, e.g. in [122]-[127], where the interval outputs are produced by extending the data or the parameters of the regression model to the interval format. However, the proposed method delivers a totally different perspective by forming a granular output space. Furthermore, the conceptual and ensuing algorithmic setting are general in the sense that it applies to a variety of categories of models including such commonly studied as neural networks, rule-based models, cognitive maps, to name a few examples.

7.1 Granular Output Space and Its Rule-Based Realization

The underlying idea introduced in this study is to develop a mechanism of allocation of information granularity across the output space. The originally constructed model, Figure 7.1, is augmented by a functional transformation module, which along with the granulation block transforms a numeric output into an information granule. The role of the transformation module is to realize a nonlinear mapping from the numeric output to the corresponding level of information granularity (ε). The function of the granulation block is on a basis of the provided level of information granularity to translate the numeric output to the information granule. In light of the interval-valued form of information granules, the numeric output y is elevated to the form of a certain interval Y distributed around of y and a length determined by the level of information granularity ε. The optimization of the allocation of information granularity implies some modifications to the transformation module. The module itself is a nonlinear mapping, which can be realized in many ways. In this study, we consider its implementation in the form of a fuzzy rule-based model.

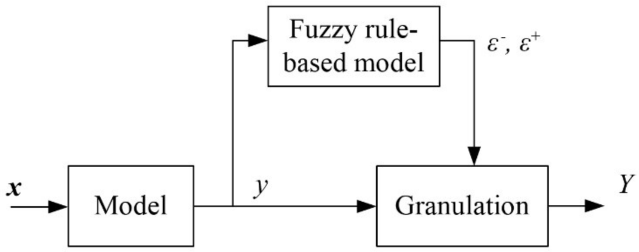


Figure 7.1. Granular model of granular output space of numeric models

The original numeric model is viewed as a multiple input - single output model M such that $M(\boldsymbol{x}) = y$. We do not make any assumption about the form of the model; it could be a neural network, regression model, cognitive map, polynomial function, etc.

The transformation module generating a level of information granularity is a rule-based model, which is described as a collection of single input – two output rules in the format

$$\text{if } y \text{ is } A_i \text{ then information granularity is } (\varepsilon_i^-, \varepsilon_i^+) \tag{7.1}$$

i=1, 2,…, c where ε_i^-, ε_i^+ imply the location of the interval of the output, with the values of ε_i^-, ε_i^+ located in the [0,1] interval while A_i is a fuzzy set formed in the output space $\boldsymbol{Y}$. From the perspective of the rule-based architecture, (1) can be viewed as a TS fuzzy rule-based model of type-0 (viz. the conclusion part is a constant). From the structural point of view, the resulting mapping is piecewise linear of the output space to the levels of information granularity.

Let us look at the above rules governing an allocation of information granularity.

(i) fuzzy sets of the condition part. Typically, as a sound initial position we consider a collection of fuzzy sets of finite supports distributed uniformly across the output space. The simplest alternative is a collection of fuzzy sets with triangular membership functions and ½ overlap between adjacent fuzzy sets, see Figure 7.2. The range of the output spreads across the extreme values of target output (namely, interval [$target_{min}$, $target_{max}$]). The fuzzy sets are uniquely described by their modal values sv_1 (= $target_{min}$), $sv_2,\ldots, sv_c$ (= $target_{max}$).

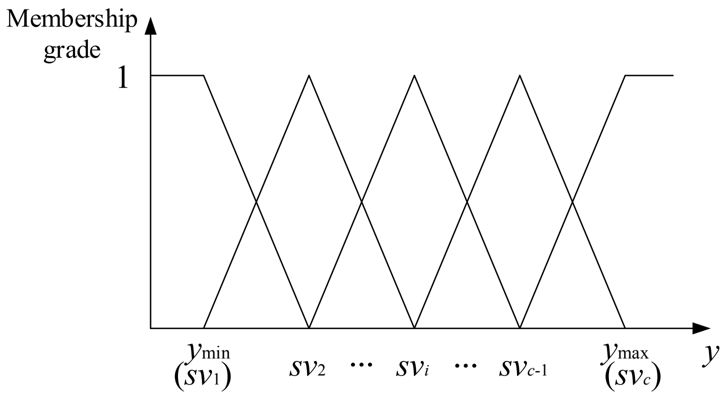


Figure 7.2. A collection of fuzzy sets defined in the output space

(ii) inference mechanism governing an allocation of information granularity. We determine the levels of information granularity (ε_i^-, ε_i^+) for any numeric output y by implementing a standard inference mechanism. As A_is form a partition (any two adjacent fuzzy sets overlap at ½ level). The allocated information granularity (ε^-, ε^+) associated with the given output y is expressed in the following way

$$\varepsilon^- = \sum_{i=1}^{c} A_i(y)\varepsilon_i^- \tag{7.2}$$

$$\varepsilon^+ = \sum_{i=1}^{c} A_i(y)\varepsilon_i^+ \tag{7.3}$$

In view of the form of the membership functions of A_i, both ε^- and ε^+ are piecewise nonlinear mappings of the output.

The information granulation module transforms the original numeric output y to an interval Y in the form,

$$Y = \left[y - \varepsilon^- range, y + \varepsilon^+ range\right] \tag{7.4}$$

where *range* = $target_{max}$-$target_{min}$. It is apparent that the length of *Y* is dependent on the specific value of *y* as ε^- and ε^+ are both functions of *y*.

7.2 Optimization of Granular Output Space

The construction of the optimal granular model through the buildup of the granular output space involves two main components: (i) allocation of information granularity through optimizing the corresponding rule-based model supporting this allocation process, and (ii) formulation of a suitable performance index using which the optimal allocation of information granularity is achieved.

Here we consider three strategies to allocate information granularity across the output space of the model (viz the allocation of levels of information granularity to each of the *c* rules).

(1) *no optimization- uniform distribution of information granularity*. In this strategy, we are not using the fuzzy rule-based model, but allocate the information granularity according to the overall level of information granularity ε, namely,

$$\varepsilon_i^- = \varepsilon_i^+ = \frac{\varepsilon}{2} \tag{7.5}$$

i=1, 2,…c. In this strategy, the specificity of the results is a linearly decreasing function of ε. We have $y_k^- = y_k - \frac{\varepsilon}{2} range \quad y_k^+ = y_k + \frac{\varepsilon}{2} range$

$$sp = \frac{1}{N}\sum_{k=1}^{N}\max\left(\left(1 - \frac{y_k^+ - y_k^-}{range}\right), 0\right) = \frac{1}{N}\sum_{k=1}^{N}\left(1 - \frac{\varepsilon \cdot range}{range}\right)$$
$$= \frac{1}{N}\sum_{k=1}^{N}(1-\varepsilon) = 1-\varepsilon$$

(2) *symmetric allocation of information granularity*. We assume the level of information granularity associated with each rule is symmetric in the following sense,

$$\varepsilon_i^- = \varepsilon_i^+ = \frac{\varepsilon_i}{2} \tag{7.6}$$

Obviously, in virtue of the constraint imposed on the level of information granularity, we have

$$\sum_{i=1}^{c} \varepsilon_i = c\varepsilon \tag{7.7}$$

We optimize the allocation of information granularity by maximizing the values of V with respect to both ε_i and the modal values of the fuzzy sets forming the condition parts of rules, namely a_2, a_3, …, a_{c-1}. The optimization problem reads as follows

$$\max_{\varepsilon_1, \varepsilon_1, \ldots, \varepsilon_c, a_2, a_3, \ldots, a_{c-1}} V$$

subject to constraint

$$\sum_{i=1}^{c} \varepsilon_i = c\varepsilon, \quad \varepsilon_i \geq 0$$

(3) *asymmetric allocation of information granularity*. Here the information granularity (ε_i^-, ε_i^+) allows the intervals to be allocated asymmetrically with respect to the original numeric value. Thus, we maximize V with respect to ε_i^-, ε_i^+ and a_2, a_3, …, a_{c-1} and the overall optimization problem reads as

$$\max_{\varepsilon_1^-, \varepsilon_2^-, \ldots, \varepsilon_c^-, \varepsilon_1^+, \varepsilon_2^+, \ldots, \varepsilon_c^+, a_2, a_3, \ldots, a_{c-1}} V$$

subject to constraint

$$\sum_{i=1}^{c} \varepsilon_i^- + \sum_{i=1}^{c} \varepsilon_i^+ = c\varepsilon, \quad \varepsilon_i^-, \varepsilon_i^+ \geq 0$$

To quantify the performance of the granular fuzzy rule-based models, we assess coverage to represent the ability of the output information granule to cover the experimental data as (2.29). However, to evaluate the specificity, we attempt another function of the length of the interval as follows,

$$sp = \frac{1}{N} \sum_{k=1}^{N} \max\left(\left(1 - \frac{y_k^+ - y_k^-}{range}\right), 0\right) \tag{7.8}$$

The maximum operation is used here to prevent the situation when y_k^+-y_k^- is higher than *range* and this could result in a negative value of the specificity. The overall performance of information granules is quantified by (2.37).

The level of information granularity is a design asset whose proper allocation is crucial to the optimization of the granular model meaning. This entails that given some level of information granularity ε, $\varepsilon \in [0, 1]$. We distribute it in such a way that the criterion (8) becomes maximize. At the same time, we request that the following balance of information granularity becomes satisfied

$$\sum_{i=1}^{c} \varepsilon_i^- + \sum_{i=1}^{c} \varepsilon_i^+ = c\varepsilon \tag{7.9}$$

To evaluate a global characterization of the model, we compute the overall performance index as described in (2.39).

7.3 Experiment Studies

A single input - single output synthetic data composed of 400 input-output pairs are described as follows

$$y = x^2 - 4 \cdot x \cdot \sin(x) \tag{7.10}$$

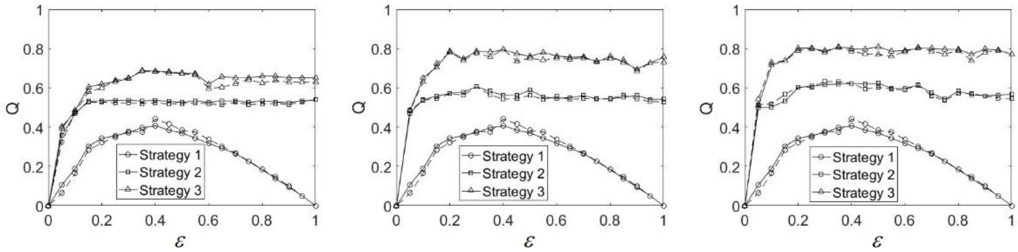
where the values of x are uniformly random distributed in the range of [0, 4]. The data is split into training set (70%) and testing set (30%).

Assume that the developed model comes in the form

$$y = x^2 \tag{7.11}$$

First, we compare the performance of the granular output allocated by the three proposed strategies (symmetrically constant allocation and fuzzy rule-based model allocation after optimization), the *V* values versus the different levels of information granularity are plotted in Figure 7.3. Here the control coefficient β is set as 1. With the first allocation strategy, Q raises at the beginning, since the coverage increases with the extending level of information granularity. Then, Q declines gradually after the coverage reaching the extreme level (in most of cases $cov = 1$) and specificity values keep going down until $sp = 0$. The improvement coming from the other two strategies is quite visible. The values of Q are not only higher but also it increases when the predetermined level of information granularity ε is lower, then it maintains at a higher level. Thus, we reckon that, with the optimization respecting to the objective function, it is helpful to make a tradeoff between the conflicting two criteria (coverage and specificity). In addition, not surprisingly, strategy 3 is better than strategy 2, because it is more flexible to optimize. The results are also confirmed in Table 7.1. The overall performance associated with the third strategy is superior to the performance produced by the two other strategies. Furthermore, the number of rules has effect on the output performance. No surprisingly, the more number of rules, the fuzzy model is more capable to capture the characteristic of the target data, so the overall performance of the granular output is better.

In Figure 7.4 and Figure 7.5, we display the details of coverage and specificity corresponding to the level of information granularity. Generally, coverage increases and specificity decreases with the increasing value of information granularity (ε). Nevertheless, due to the optimization procedures in the strategies 2 and 3, the coverage improves faster and the specificity is able to retain at a certain level. Particularly, the specificity results of the two strategies are relatively similar, which indicates that the optimization regarding to the objective function is conducive to maintain the specificity of the granular output. Meanwhile, the symmetric and asymmetric allocation strategy have different impact on coverage performance. In strategy 3, the coverage can quickly reach 100%, while in strategy 2, the coverage never approaches to 100% since it has to make tradeoff with another conflicting criterion (specificity). In conclusion, the asymmetric allocation strategy should be the more reasonable option.

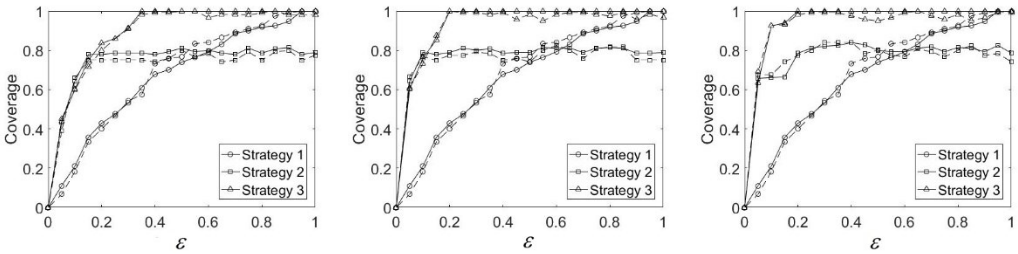


(a) (b) (c)

Figure 7.3. Plots of the fitness values (Q) vs. the level of information granularity (ε): (a) c = 5, (b) c = 10, (c) c = 20. solid lines: training data set, dashed lines: testing data set.

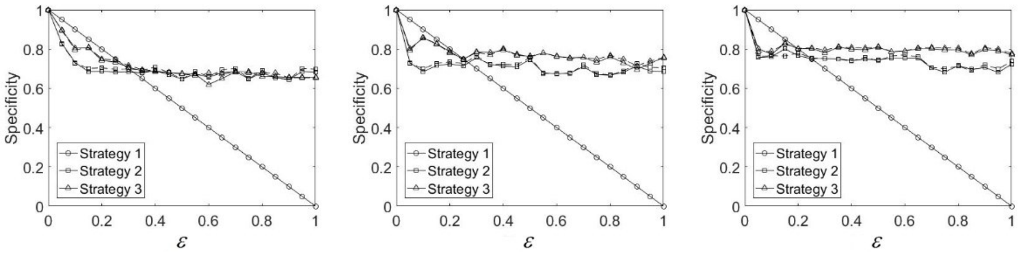
Table 7.1. V values of granular models' outputs

Data set		c = 5	c = 10	c = 20
Strategy 1	Training	0.2447	0.2447	0.2447
	Testing	0.2460	0.2460	0.2460
Strategy 2	Training	0.4955	0.5290	0.5527
	Testing	0.4857	0.5219	0.5512
Strategy 3	Training	0.5581	0.7015	**0.7365**
	Testing	0.5590	0.6922	**0.7289**



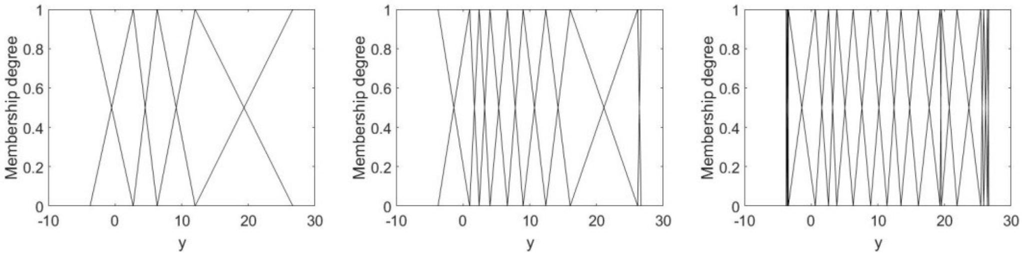
(a) (b) (c)

Figure 7.4. Plots of the Coverage vs. the level of information granularity (ε): (a) c = 5, (b) c = 10, (c) c = 20. Solid lines: Training data set, Dashed lines: Testing data set.



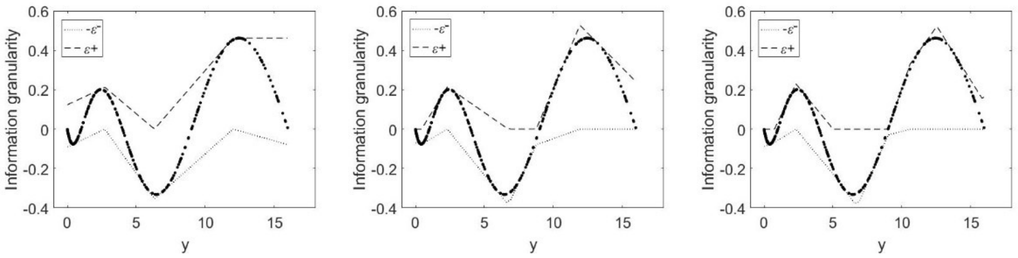
(a) (b) (c)

Figure 7.5. Plots of Specificity vs. the level of information granularity (ε): (a) c = 5, (b) c = 10, (c) c = 20. Solid lines: Training data set, Dashed lines: Testing data set.



(a) (b) (c)

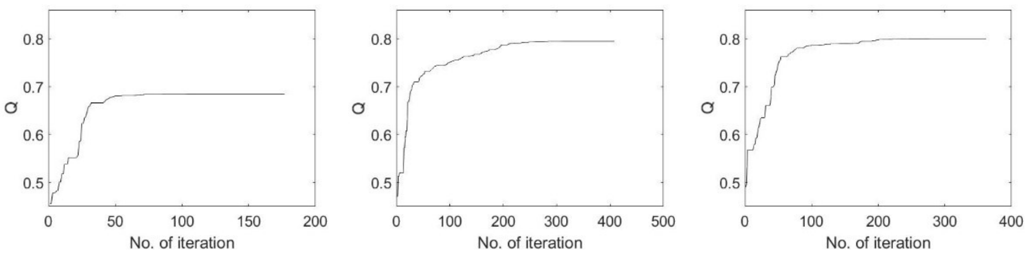
Figure 7.6. Distributions of the optimized fuzzy sets of the fuzzy models (ε = 0.4): (a) c = 5, (b) c = 10, (c) c = 20.



(a) (b) (c)

Figure 7.7. Distribution of allocated information granularity (ε = 0.4): (a) c = 5, (b) c = 10, (c) c = 20. points: *corresponding error*.

To demonstrate some detail of the allocation of information granularity obtained by the third strategy, we report some more results taking that the predetermined level of information granularity is 0.4 as an example. In Figure 7.6, we display the distribution of the optimized fuzzy numbers (sv_i). The change of the distributions after optimization is quite visible. Then, in Figure 7.7 we illustrate the distribution of the allocated information granularity ($-\varepsilon_i^-$, ε_i^+). The dots in the plots are shown the *corresponding error* of the model outputs, which is calculated as (*target* - *y*)/*range*. As we can see from the figures, the distribution is related to the error of target output. Moreover, the more number of rules, the more elaborated the distribution is.

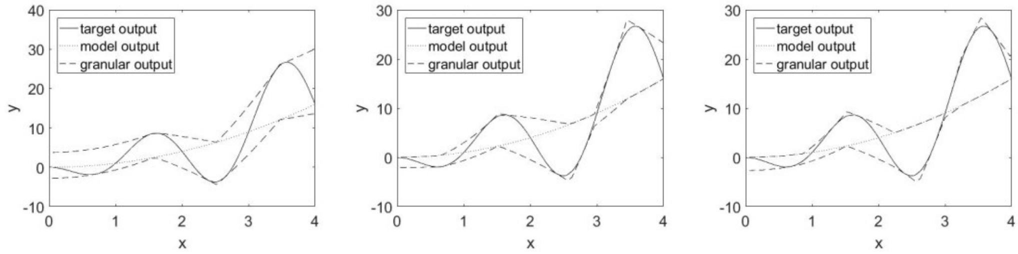


(a) (b) (c)

Figure 7.8. Plots of the performance of the fitness function in successive generations (ε = 0.4): (a) c = 5, (b) c = 10, (c) c = 20.

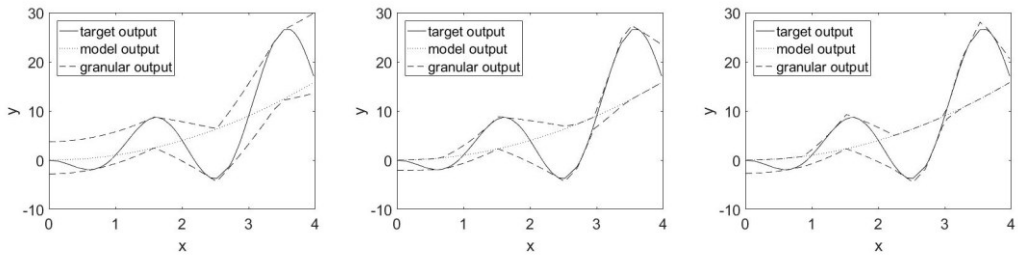
The optimization processes are included in Figure 7.8. It is noticeable that the most visible improvement occurs at some initial generations of the method and then beyond some certain generations (60 ~ 200) there is no increase in the values of the fitness function. To illustrate an overall impression of the granular outputs and numeric output, in Figure 7.9 and Figure 7.10, the target output, model output and granular outputs are plotted. As it is seen, even though the model output is inaccuracy, the granular output covers the most of target output and captures its main characteristic. Furthermore, the number of rules of the fuzzy rule-based model for allocation has significant effect on the granular output. Generally, the more number of rules, the more specific the granular output is.

In Figure 7.11 and Figure 7.12, we contrast 3 different values of the control coefficient β to demonstrate its impact on the output performance. As shown in the figures, if β is 0, the fitness function is only focused on coverage criterion, so the granular intervals are too extended. If β equals 3, the specificity has too much influence on the overall performance to have excepted coverage performance. By contrast, the outcomes are much better when β is set as 0.5 or 1. In conclusion, it is reasonable to set the control coefficient β around 1 (probably 1.0±0.5).



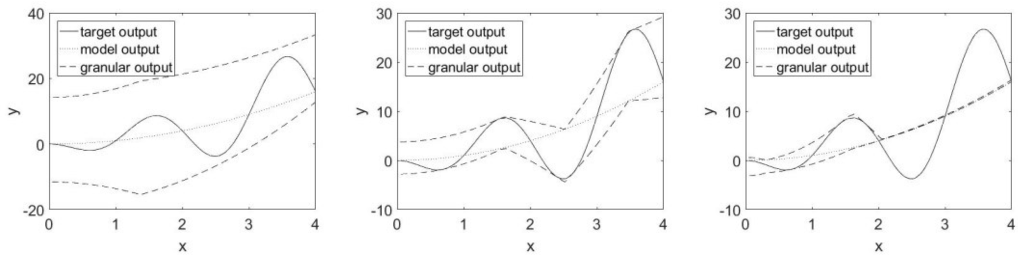
(a) (b) (c)

Figure 7.9. Plots of target, model and granular output of training data sets (ε = 0.4): (a) c = 5, (b) c = 10, (c) c = 20.



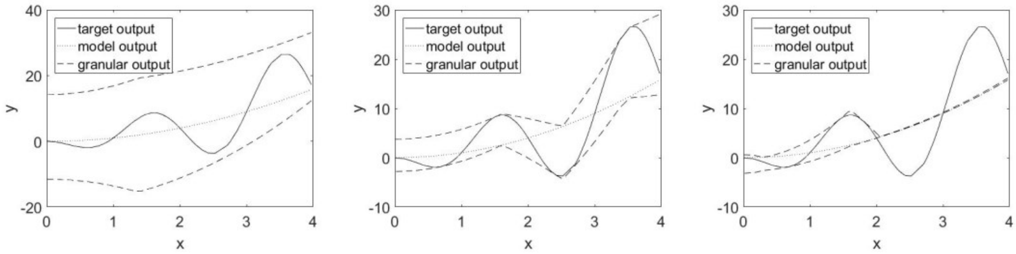
(a) (b) (c)

Figure 7.10. Plots of target, model and granular output of testing data sets (ε = 0.4): (a) c = 5, (b) c = 10, (c) c = 20.



(a) (b) (c)

Figure 7.11. Plots of target, model and granular output of training data sets (ε = 0.4, c = 5): (a) β = 0, (b) β = 0.5, (c) β = 3.



(a) (b) (c)

Figure 7.12. Plots of target, model and granular output of testing data sets (ε = 0.4, c = 5): (a) β = 0, (b) β = 0.5, (c) β = 3.

7.4 Summary

The study offered the way of designing granular models through the development of the granular output space. The development of the model is composed of the two essential phases. The first one concerns a formation of the numeric model while in the second phase the model becomes endowed with a granular output space. The optimization of the space engages information granularity being viewed as the design asset whose optimal allocation helps maximize the performance index capturing the quality of information granules formed by the model, namely the coverage and the specificity factor. In contrast to the formation of the granular parameter space studied so far (completed either through allocation of information granularity across parameters of the model or the input space), the alternative presented in this study comes with competitive results and associates with a lower computing overhead.

Chapter 8

Major contributions, conclusions and Future Studies

This thesis focuses on the development and evaluation of granular fuzzy rule-based models to offer a new and attractive system-modeling methodology at the level of information granules.

8.1 Major contributions

The methods proposed in this thesis revolve around the main procedures and components of a fuzzy rule-based model framework to establish a higher-level granular form. The major contributions proposed in this dissertation are summarized as follows,

(1) We have augmented the granulation-degranulation scheme, which is beneficial to reduce the deterioration of the reconstruction results and to enhance the performance of the overall granulation-degranulation scheme. The improved granulation-degranulation scheme is meaningful for generating better clusters (prototypes) for building fuzzy rule-based models.

(2) We consider the process between the information granules by using logic operators. In Granular Computing, there is also a need for aggregating several information granules into a single output granule. Therefore, we present a study to characterize them by analytical properties. This study is helpful for selecting an appropriate logic operator for fuzzy modeling.

(3) We consider three main functional modules of fuzzy rule-based models to develop them to granular forms. First, we focus on the concept of granular input space. The direct linkages have been identified between the granular space formed in this way and the analysis of the impact of input variables in the already developed models. Second, we design granular fuzzy rule-based models by granulating several components of processing modules of fuzzy models such as the interactive rule matrix, the prototypes and parameters in the condition parts of rules, and the parameters of the conclusion parts of rules. Finally, we

offer a way to design granular models through the development of granular output space.

8.2 Conclusions

Overall, the studies discussed in this dissertation show some interesting conclusions:

(1) The experimental studies demonstrate that the resulting granular outputs could cover most of target evidences as specifically as possible, which makes the granular output of the fuzzy rule-based model more tolerant to error and of higher practical relevance.

(2) We show that the coverage and specificity measures serve as the two essential measures —the coverage and specificity— that quantify a well-rounded way of expressing the quality of the granular model. The proposed granular fuzzy rule-based models are evaluated and analyzed by invoking two criteria guided by principle of justifiable information granularity. Basically, the increase in specificity comes at the expense of the decreasing values of coverage, which is realized as the conflicting nature of the coverage and specificity criteria. The characterization of the granular model in terms of its coverage-specificity relationships—or the global descriptor coming in the form of the AUC (V) measure—becomes beneficial to a holistic assessment of the quality of the rule-based models. These two components are crucial in the evaluation of the performance of the granular fuzzy rule-based models.

(3) To optimize the allocation of the information granularity of granular fuzzy rule-based models, we usually tested and implemented several scenarios: for instance, different number of prototypes (rules) and uniform and non-uniform allocation strategies. Generally, the fuzzy rule-based models constructed with more rules perform better, as they are better able to capture the data structure. The non-uniform strategy often provides better performance because of its flexibility in the allocation of information granularity. The complexity of

models and allocation strategies is related to both performance and efficiency; therefore, we need to have the trade-off in mind for setting the models.

(4) The optimizations we carried out in the experiments implement PSO, DE and their variants, because in most cases the fitness function is nonconvex or not deferentially guaranteed. In the experimental studies, these population-based search algorithms show robustness, performance, and implementation flexibility for the allocation of information granularity.

8.3 Future Studies

Several promising directions are worth placing on the research agenda of future research. We may highlight three of them by exhibiting some potential and direct practical implications:

(1) In the realm of reconstruction of information granules, First, various alternatives of interaction matrices could be considered, especially for highly dimensional problems. Second, one could consider this approach as a certain way of coping with a non-stationary environment (say, data streams) and in this setting, the introduced adjustment mechanisms endow the method with the required calibration capabilities. The quality of the clusters produced for some initial data in the stream could be improved not by running clustering on a new data but rather by adjusting the existing structure in the way discussed above. In this sense, the clusters are built in an evolutionary manner rather than in an abrupt way when moving from one segment of the data stream to another and in this way, retain a desirable property of continuity.

(2) Engagement of other formalisms of Granular Computing. The principles of granular fuzzy models were outlined with the use of intervals or fuzzy sets. While this was done for illustrative purposes as being conceptually the simplest and computationally feasible, the fundamentals can be used exploited by involving other formal settings.

(3) Designing of granular models of higher type. Following the design process discussed in this study, the numeric parameters of the model are transformed (generalized) into information granules of Type-1. To enhance the coverage of the data, the granular parameters and constructions can be further generalized to build information granules of Type-2 (say, granular intervals generalizing intervals with numeric bounds to the intervals whose bounds are information granules themselves).

(4) The concept of granularity can be contrasted with the ideas of granular input, parameters, and output space in terms of their performances. Along the same line, a combined topology of granular multiple modules can be considered and optimized, which comes from a hybrid approach of granular modeling.

(5) Two criteria are used to evaluate the performance of a granular fuzzy rule-based models: coverage and specificity. We are aware of the conflicting nature of these two indices, but we concentrated on the construction of the granular fuzzy modeling in previous studies and followed a relatively simplified strategy: We calculated the AUC of the two indices or multiply them into a global indicator for the optimization. The weakness of this process is obvious, so a viable alternative is to invoke bi-objective optimization with the intent to establish some sound trade-offs in a family of solutions.

Bibliography

[1] LA Zadeh. "Fuzzy sets." *Information and control* 8.3 (1965): 338-353.

[2] LA Zadeh. "Fuzzy logic—a personal perspective." *Fuzzy Sets and Systems* 281 (2015): 4-20.

[3] T Takagi, M Sugeno. "Fuzzy identification of systems and its applications to modeling and control." *Systems, IEEE Transactions on Man and Cybernetics* 1 (1985): 116-132.

[4] EH Mamdani, S Assilian. "An experiment in linguistic synthesis with a fuzzy logic controller." *International journal of man-machine studies* 7.1 (1975): 1-13.

[5] W Pedrycz, M Reformat. "Rule-based modeling of nonlinear relationships." *IEEE Transactions on Fuzzy Systems* 5.2 (1997): 256-269.

[6] W Pedrycz, M Reformat. "Evolutionary fuzzy modeling." *IEEE Transactions on Fuzzy Systems* 11.5 (2003): 652-665.

[7] B Rezaee, MHF Zarandi. "Data-driven fuzzy modeling for Takagi–Sugeno–Kang fuzzy system." *Information Sciences* 180.2 (2010): 241-255.

[8] Q Zhang, M Mahfouf. "A hierarchical Mamdani-type fuzzy modelling approach with new training data selection and multi-objective optimisation mechanisms: A special application for the prediction of mechanical properties of alloy steels." *Applied soft computing* 11.2 (2011): 2419-2443.

[9] P Baranyi. "The Generalized TP model transformation for T–S fuzzy model manipulation and generalized stability verification." *IEEE Transactions on Fuzzy Systems* 22.4 (2014): 934-948.

[10] Z Deng, Y Jiang, KS Choi, FL Chung. "Knowledge-leverage-based TSK fuzzy system modeling." *Neural Networks and Learning Systems, IEEE Transactions on* 24.8 (2013): 1200-1212.

[11] W Pedrycz, F Gomide. *Fuzzy systems engineering: toward human-centric computing*. John Wiley & Sons, 2007.

[12] W Pedrycz. *Granular computing: analysis and design of intelligent systems*. CRC press, 2013.

[13] LA Zadeh. "Fuzzy sets and information granularity." *Advances in fuzzy set theory and applications* 11 (1979): 3-18.

[14] LA Zadeh. "Fuzzy graphs, rough sets and information granularity." *Proceedings of the Third International Workshop on Rough Sets and Soft Computings*, San Jose. 1994.

[15] Y Qian, J Liang, W Wu, C Dang. "Information granularity in fuzzy binary GrC model." *IEEE Transactions on Fuzzy Systems* 19.2 (2011): 253-264.

[16] W Pedrycz, G Succi, A Sillitti, J Iljazi. "Data description: a general framework of information granules." *Knowledge-Based Systems* 80 (2015): 98-108.

[17] W Pedrycz, M Song. "Analytic hierarchy process (AHP) in group decision making and its optimization with an allocation of information granularity." *IEEE Transactions on Fuzzy Systems* 19.3 (2011): 527-539.

[18] Y Yao, L Zhao. "A measurement theory view on the granularity of partitions." *Information Sciences* 213 (2012): 1-13.

[19] W Lu, W Pedrycz, X Liu, J Yang, P Li. "The modeling of time series based on fuzzy information granules." *Expert Systems with Applications* 41.8 (2014): 3799-3808.

[20] H Zhu, S Ding, L Xu, L Zhang. "Research and development of granularity clustering." *Computer Science for Environmental Engineering and EcoInformatics*. Springer Berlin Heidelberg, 2011. 253-258.

[21] A Gacek. "From clustering to granular clustering: a granular representation of data in pattern recognition and system modeling." *IFSA World Congress and NAFIPS Annual Meeting* (IFSA/NAFIPS), 2013 Joint. IEEE, 2013.

[22] MA Sanchez, O Castillo, JR Castro, P Melin. "Fuzzy granular gravitational clustering algorithm for multivariate data." *Information Sciences* 279 (2014): 498-511.

[23] LA Zadeh. "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic." *Fuzzy sets and systems* 90.2 (1997): 111-127.

[24] A Bargiela, W Pedrycz. *Granular computing: an introduction*. Vol. 717. Springer Science & Business Media, 2012.

[25] A. Gacek. "Granular modelling of signals: a framework of granular computing." *Information Sciences* 221 (2013): 1-11.

[26] VJ Kok, CS Chan. "GrCS: Granular Computing-Based Crowd Segmentation." *IEEE transactions on cybernetics* 47.5 (2017): 1157-1168.

[27] P Hońko. "Properties of a Granular Computing Framework for Mining Relational Data." *International Journal of Intelligent Systems* 32.3 (2017): 227-248.

[28] JT Yao, AV Vasilakos, W Pedrycz. "Granular computing: perspectives and challenges." *IEEE Transactions on Cybernetics* 43.6 (2013): 1977-1989.

[29] A Bargiela, W Pedrycz. "Toward a theory of granular computing for human-centered information processing." *IEEE Transactions on Fuzzy Systems* 16.2 (2008): 320-330.

[30] W Pedrycz, R Al-Hmouz, AS Balamash. "Designing granular fuzzy models: A hierarchical approach to fuzzy modeling." *Knowledge-Based Systems* 76 (2015): 42-52.

[31] W Lu, L Zhang, W Pedrycz, J Yang, X Liu. "The granular extension of Sugeno-type fuzzy models based on optimal allocation of information granularity and its application to forecasting of time series." *Applied Soft Computing* 42 (2016): 38-52.

[32] W Pedrycz, R Al-Hmouz, AS Balamash, A Morfeq. "Modeling with linguistic entities and linguistic descriptors: a perspective of granular computing." *Soft Computing* 21.7 (2017): 1833-1845.

[33] W Pedrycz, M Song. "Granular fuzzy models: a study in knowledge management in fuzzy modeling." *International Journal of Approximate Reasoning* 53.7 (2012): 1061-1079.

[34] W Pedrycz, et al. "The design of free structure granular mappings: the use of the principle of justifiable granularity." *IEEE Transactions on Cybernetics* 43.6 (2013): 2105-2113.

[35] W Pedrycz. "From numeric to granular description and interpretation of information granules." *Fundamenta Informaticae* 127.1-4 (2013): 399-412.

[36] OF Reyes-Galaviz, W Pedrycz. "Granular fuzzy modeling with evolving hyperboxes in multi-dimensional space of numerical data." *Neurocomputing* 168 (2015): 240-253.

[37] MM Ahmed, NAM Isa. "Information granularity model for evolving context-based fuzzy system." *Applied Soft Computing* 33 (2015): 183-196.

[38] W Pedrycz. "From Fuzzy Models to Granular Fuzzy Models." *International Journal of Computational Intelligence Systems* 9.sup1 (2016): 35-42

[39] M Song, Y Wang. "Human centricity and information granularity in the agenda of theories and applications of soft computing." *Applied Soft Computing* 27 (2015): 610-613.

[40] Z Pawlak. "Rough sets." *International Journal of Parallel Programming* 11.5 (1982): 341-356.

[41] W Pedrycz. "Shadowed sets: representing and processing fuzzy sets." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 28.1 (1998): 103-109.

[42] OF Reyes-Galaviz, W Pedrycz. "Granular fuzzy models: Analysis, design, and evaluation." *International Journal of Approximate Reasoning* 64 (2015): 1-19.

[43] W Pedrycz. "From fuzzy data analysis and fuzzy regression to granular fuzzy data analysis." *Fuzzy Sets and Systems* 274 (2015): 12-17.

[44] W Pedrycz. "The principle of justifiable granularity and an optimization of information granularity allocation as fundamentals of granular computing." *Journal of Information Processing Systems* 7.3 (2011): 397-412.

[45] X Wang, W Pedrycz, A Gacek, X Liu. "From numeric data to information granules: A design through clustering and the principle of justifiable granularity." *Knowledge-Based Systems* 101 (2016): 100-113.

[46] D Wang, W Pedrycz, Z Li. "Design of granular interval-valued information granules with the use of the principle of justifiable granularity and their

applications to system modeling of higher type." *Soft Computing* 20.6 (2016): 2119-2134.

[47] W Pedrycz, X Wang. "Designing Fuzzy Sets with the Use of the Parametric Principle of Justifiable Granularity." *IEEE Transactions on Fuzzy Systems* 24.2 (2016): 489-496.

[48] X Zhu, W Pedrycz, Z Li. "Granular data description: designing ellipsoidal information granules." *IEEE transactions on cybernetics* (2017).

[49] W Pedrycz, B Russo, G Succi. "Knowledge transfer in system modeling and its realization through an optimal allocation of information granularity." *Applied Soft Computing* 12.8 (2012): 1985-1995.

[50] FJ Cabrerizo, R Ureña, W Pedrycz. "Building consensus in group decision making with an allocation of information granularity." *Fuzzy Sets and Systems* 255 (2014): 115-127.

[51] W Pedrycz. "From fuzzy sets to shadowed sets: interpretation and computing." *International Journal of Intelligent Systems* 24.1 (2009): 48-61.

[52] O Cordón. "A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: Designing interpretable genetic fuzzy systems." *International Journal of Approximate Reasoning* 52.6 (2011): 894-913.

[53] J C Bezdek, R Ehrlich, W Full. "FCM: The fuzzy c-means clustering algorithm." *Computers & Geosciences* 10.2 (1984): 191-203.

[54] S R Kannan, S Ramathilagam, PC Chung. "Effective fuzzy c-means clustering algorithms for data clustering problems." *Expert Systems with Applications* 39.7 (2012): 6292-6300.

[55] TC Havens, JC Bezdek, C Leckie. "Fuzzy c-means algorithms for very large data." *IEEE Transactions on Fuzzy Systems* 20.6 (2012): 1130-1146.

[56] T Li, L Zhang, W Lu, H Hou, X Liu, W Pedrycz, C Zhong. "Interval Kernel Fuzzy C-Means Clustering of Incomplete Data." *Neurocomputing* (2017).

[57] S Askari, N Montazerin, MHF Zarandi. "Generalized Possibilistic Fuzzy C-Means with novel cluster validity indices for clustering noisy data." *Applied Soft Computing* 53 (2017): 262-283.

[58] CL Tang, SG Wang, and W Xu. "New fuzzy c-means clustering model based on the data weighted approach." *Data & Knowledge Engineering* 69.9 (2010): 881-900.

[59] D Graves, and W Pedrycz. "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study." *Fuzzy sets and systems* 161.4 (2010): 522-543.

[60] G Tsekouras, H Sarimveis, E Kavakli, G Bafas. "A hierarchical fuzzy-clustering approach to fuzzy modeling." *Fuzzy Sets and Systems* 150.2 (2005): 245-266.

[61] M Chadli, HR Karimi. "Robust observer design for unknown inputs Takagi–Sugeno models." *IEEE Transactions on Fuzzy Systems* 21.1 (2013): 158-164.

[62] B Ding, H Pan. "Dynamic Output Feedback-Predictive Control of a Takagi–Sugeno Model with Bounded Disturbance." *IEEE Transactions on Fuzzy Systems* 25.3 (2017): 653-667.

[63] P Hušek, K Narenathreyas. "Aircraft longitudinal motion control based on Takagi–Sugeno fuzzy model." *Applied Soft Computing* 49 (2016): 269-278.

[64] A Almaksour, E Anquetil. "ILClass: Error-driven antecedent learning for evolving Takagi-Sugeno classification systems." *Applied Soft Computing* 19 (2014): 419-429.

[65] H S Park, W Pedrycz, Sung-Kwun Oh. "Granular neural networks and their development through context-based clustering and adjustable dimensionality of receptive fields." *IEEE Transactions on Neural Networks* 20.10 (2009): 1604-1616.

[66] W Pedrycz, H Izakian. "Cluster-centric fuzzy modeling." *IEEE transactions on fuzzy systems* 22.6 (2014): 1585-1597.

[67] M. Delgado, A. F. Gómez-Skarmeta. "A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling." *IEEE Transactions on Fuzzy Systems* 5.2 (1997): 223-233.

[68] J Qiao, W Li, XJ Zeng, H Han. "Identification of fuzzy neural networks by forward recursive input-output clustering and accurate similarity analysis." *Applied Soft Computing* 49 (2016): 524-543.

[69] S Dowdy, S Wearden, D Chilko. *Statistics for Research*, John Wiley & Sons, 2011.

[70] M Hollander, DA Wolfe, E Chicken. *Nonparametric Statistical Methods*. John Wiley & Sons, 2013.

[71] E Zeidler, W Hackbusch, HR Schwarz, B Hunt. *Oxford users' guide to mathematics*. Oxford University Press, 2004.

[72] MP Fay, MA Proschan. "Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules." *Statistics surveys* 4 (2010): 1.

[73] J Kennedy. "*Particle swarm optimization*." Encyclopedia of Machine Learning. Springer US, 2010. 760-766.

[74] R Storn, K Price. "Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces." *Journal of global optimization* 11.4 (1997): 341-359.

[75] OF Reyes-Galaviz, W Pedrycz. Granular fuzzy modeling with evolving hyperboxes in multi-dimensional space of numerical data. *Neurocomputing* 168 (2015): 240-253.

[76] A Balamash, W Pedrycz, R Al-Hmouz. "An expansion of fuzzy information granules through successive refinements of their information content and their use to system modeling." *Expert Systems with Applications* 42.6 (2015): 2985-2997.

[77] MS Yang. "A survey of fuzzy clustering." *Mathematical and Computer modelling* 18.11 (1993): 1-16.

[78] W Pedrycz, A Gacek, X Wang. "Clustering in augmented space of granular constraints: a study in knowledge-based clustering." *Pattern Recognition Letters* 67 (2015): 122-129.

[79] OF Reyes-Galaviz, W Pedrycz. "Granular fuzzy models: analysis, design, and evaluation." *International Journal of Approximate Reasoning* 64 (2015): 1-19.

[80] W Pedrycz, G Succi, A Sillitti, J Iljazi "Data description: a general framework of information granules." *Knowledge-Based Systems* 80 (2015): 98-108.

[81] W Pedrycz, JVD Oliveira. "A development of fuzzy encoding and decoding through fuzzy clustering." *IEEE Transactions on Instrumentation and Measurement* 57.4 (2008): 829-837.

[82] A Gacek, W Pedrycz. "Clustering granular data and their characterization with information granules of higher type." *IEEE Transactions on Fuzzy Systems* 23.4 (2015): 850-860.

[83] X Zhu, W Pedrycz, Z Li. "Granular Encoders and Decoders: A Study in Processing Information Granules." *IEEE Transactions on Fuzzy Systems* (2016).

[84] H Izakian, W Pedrycz. Anomaly detection and characterization in spatial time series data: A cluster-centric approach. *IEEE Transactions on Fuzzy Systems* 22.6 (2014): 1612-1624.

[85] OF Reyes-Galaviz, W Pedrycz. Enhancement of the classification and reconstruction performance of Fuzzy C-Means with refinements of prototypes. *Fuzzy Sets and Systems* 318 (2017): 80-99.

[86] J. Zhang, AC. Sanderson. "JADE: adaptive differential evolution with optional external archive." *IEEE Transactions on evolutionary computation* 13.5 (2009): 945-958.

[87] Tanabe, Ryo, A. S. Fukunaga. Success-history based parameter adaptation for differential evolution. *Evolutionary Computation (CEC), 2013 IEEE Congress on*. *IEEE*, 2013.

[88] Tanabe, Ryo, A. S. Fukunaga. Improving the search performance of SHADE using linear population size reduction. *Evolutionary Computation (CEC), 2014 IEEE Congress on. IEEE*, 2014.

[89] J Drewniak, P Drygaś, E Rak. "Distributivity between uninorms and nullnorms." *Fuzzy Sets and Systems* 159.13 (2008): 1646-1657.

[90] EP Klement, R Mesiar, E. Pap. *Triangular Norms*, Springer, 2000.

[91] E Marchioni, F Montagna. "On triangular norms and uninorms definable in." *International Journal of Approximate Reasoning* 47.2 (2008): 179-201.

[92] J Dombi. "Towards a general class of operators for fuzzy systems." *IEEE Transactions on Fuzzy Systems* 16.2 (2008): 477-484.

[93] M Mas, M Monserrat, D Ruiz-Aguilera, J Torrens. "Migrative uninorms and nullnorms over t-norms and t-conorms." *Fuzzy Sets and Systems* 261 (2015): 20-32.

[94] A Xie, H Liu. "On the distributivity of uninorms over nullnorms." *Fuzzy Sets and Systems* 211 (2013): 62-72.

[95] M Mas, M Monserrat, J Torrens. "A survey on fuzzy implication functions." *IEEE Transactions on fuzzy systems* 15.6 (2007): 1107-1121.

[96] D Pei, R Yang. "Hierarchical structure and applications of fuzzy logical systems." *International Journal of Approximate Reasoning* 54.9 (2013): 1483-1495.

[97] M Zheng, Z Shi, Y Liu. "Triple I method of approximate reasoning on Atanassov's intuitionistic fuzzy sets." *International Journal of Approximate Reasoning* 55.6 (2014): 1369-1382.

[98] Y Tang, X Yang. "Symmetric implicational method of fuzzy reasoning." *International Journal of Approximate Reasoning* 54.8 (2013): 1034-1048.

[99] L Troiano, LJ Rodríguez-Muñiz. "A statistical study of differences and similarities among aggregation functions." *Logic Journal of IGPL* 19.2 (2011): 415-424.

[100] L Troiano, LJ Rodríguez-Muñiz, P Marinaro, I Díaz. "Statistical analysis of parametric t-norms." *Information Sciences* 257 (2014): 138-162.

[101] W Pedrycz, A Bargiela. "An optimization of allocation of information granularity in the interpretation of data structures: toward granular fuzzy clustering." *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 42.3 (2012): 582-590.

[102] A Bargiela, W Pedrycz. "A model of granular data: a design problem with the Tchebyschev FCM." *Soft Computing* 9.3 (2005): 155-163.

[103] JJ Liang, AK Qin, PN Suganthan. "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions." *IEEE transactions on evolutionary computation* 10.3 (2006): 281-295.

[104] W Tian. "A review of sensitivity analysis methods in building energy analysis." *Renewable and Sustainable Energy Reviews* 20 (2013): 411-419.

[105] B Iooss, and M Ribatet. "Global sensitivity analysis of computer models with functional inputs." *Reliability Engineering & System Safety* 94.7 (2009): 1194-1204.

[106] A Saltelli, M Ratto, T Andres, F Campolongo, J Cariboni. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[107] F Cannavó. "Sensitivity analysis for volcanic source modeling quality assessment and model selection." *Computers & Geosciences* 44 (2012): 52-59.

[108] A Saltelli, P Annoni, I Azzini, F Campolongo. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." *Computer Physics Communications* 181.2 (2010): 259-270.

[109] M Delgado, N Marín, D Sánchez. "Fuzzy association rules: general model and applications." *IEEE Transactions on Fuzzy Systems* 11.2 (2003): 214-225.

[110] CH Weng, YL Chen. "Mining fuzzy association rules from uncertain data." *Knowledge and Information Systems* 23.2 (2010): 129-152.

[111] J Rauch. "Logic of association rules." *Applied Intelligence* 22.1 (2005): 9-28.

[112] B Pei, S Zhao, H Chen, X Zhou, D Chen. "FARP: Mining fuzzy association rules from a probabilistic quantitative database." *Information Sciences* 237 (2013): 242-260.

[113] LG Zhou, HY Chen. "Continuous generalized OWA operator and its application to decision making." *Fuzzy Sets and Systems* 168.1 (2011): 18-34.

[114] OE PavlačKa. "Modeling uncertain variables of the weighted average operation by fuzzy vectors." *Information Sciences* 181.22 (2011): 4969-4992.

[115] SM Chen, YC Chang. "Weighted fuzzy interpolative reasoning for sparse fuzzy rule-based systems." *Expert Systems with Applications* 38.8 (2011): 9564-9572.

[116] S Garcia-Jimenez, H Bustince. "Overlap Indices: Construction of and Application to Interpolative Fuzzy Systems." *IEEE Transactions on Fuzzy Systems* 23.4 (2015): 1259-1273.

[117] W Wong, M Sun. "Deep learning regularized Fisher mappings." *IEEE Transactions on Neural Networks* 22.10 (2011): 1668-1675.

[118] R Salakhutdinov, JB Tenenbaum, A Torralba. "Learning with hierarchical-deep models." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.8 (2013): 1958-1971.

[119] A Khosravi, S Nahavandi, D Creighton, "Construction of optimal prediction intervals for load forecasting problems, "*IEEE Transactions on Power Systems*, 25, 2010, 1496-1503.

[120] DL Shrestha, DP Solomatine, "Machine learning approaches for estimation of prediction interval for the model output, " *Neural Networks*, vol. 19, 2006, 225-235.

[121] M Fazzolari, R Alcala, Y Nojima. "A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions." *IEEE Transactions on Fuzzy Systems* 21.1 (2013): 45-65.

[122] M Hladík, M Černý. "Interval regression by tolerance analysis approach." *Fuzzy Sets and Systems* 193 (2012): 85-107.

[123] H Tanaka, H Lee. "Interval regression analysis by quadratic programming approach." *IEEE Transactions on Fuzzy Systems* 6.4 (1998): 473-481.

[124] M Černý, J Antoch, M Hladík. "On the possibilistic approach to linear regression models involving uncertain, indeterminate or interval data." *Information Sciences* 244 (2013): 26-47.

[125] A Blanco-Fernández, A Colubi, M García-Bárzana. "A set arithmetic-based linear regression model for modelling interval-valued responses through real-valued variables." *Information Sciences* 247 (2013): 109-122.

[126] S Dias, P Brito. "Off the beaten track: A new linear model for interval data." *European Journal of Operational Research* 258.3 (2017): 1118-1130.

[127] A Wójtowicz, P Żywica, A Stachowiak. "Solving the problem of incomplete data in medical diagnosis via interval modeling." *Applied Soft Computing* 47 (2016): 424-437.

[128] X Hu, W Pedrycz, G Wu, X Wang. "Data reconstruction with information granules: An augmented method of fuzzy clustering." *Applied Soft Computing* 55 (2017): 523-532.

[129] X Hu, W Pedrycz, X Wang. "Comparative analysis of logic operators: a perspective of statistical testing and granular computing." *International Journal of Approximate Reasoning* 66 (2015): 73-90.

[130] X Hu, W Pedrycz, X Wang. "Optimal allocation of information granularity in system modeling through the maximization of information specificity: A development of granular input space." *Applied Soft Computing* 42 (2016): 410-422.

[131] X Hu, W Pedrycz, O Castillo, P Melin. "Fuzzy rule-based models with interactive rules and their granular generalization." *Fuzzy Sets and Systems* 307 (2017): 1-28.

[132] X Hu, W Pedrycz, X Wang. "Granular Fuzzy Rule-based Models: A Study in a Comprehensive Evaluation of Fuzzy Models." *IEEE Transactions on Fuzzy Systems* (2016). DOI: 10.1109/TFUZZ.2016.2612300.

[133] X Hu, W Pedrycz, X Wang. "From fuzzy rule-based models to their granular generalizations." *Knowledge-Based Systems* 124 (2017): 133-143.

[134] X Hu, W Pedrycz, X Wang. "Development of Granular Models Through the Design a Granular Output Spaces." *Knowledge-Based Systems* 134 (2017): 159–171.