

# SELF-LEARNING AND ADAPTIVE ALGORITHMS FOR BUSINESS APPLICATIONS

This page intentionally left blank

# SELF-LEARNING AND ADAPTIVE ALGORITHMS FOR BUSINESS APPLICATIONS

A Guide to Adaptive Neuro-fuzzy  
Systems for Fuzzy Clustering under  
Uncertainty Conditions

BY

**ZHENGBING HU**

*Central China Normal University, China*

**YEVGENIY V. BODYANSKIY**

*Kharkiv National University of Radio Electronics, Ukraine*

**OLEKSII K. TYSHCHENKO**

*University of Ostrava, Czech Republic*



United Kingdom — North America — Japan  
India — Malaysia — China

Emerald Publishing Limited  
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2019

Copyright © 2019 Emerald Publishing Limited

**Reprints and permissions service**

Contact: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

**British Library Cataloguing in Publication Data**

A catalogue record for this book is available from the British Library

ISBN: 978-1-83867-174-7 (Print)

ISBN: 978-1-83867-171-6 (Online)

ISBN: 978-1-83867-173-0 (Epub)



Certificate Number 1985  
ISO 14001

ISOQAR certified  
Management System,  
awarded to Emerald  
for adherence to  
Environmental  
standard  
ISO 14001:2004.



INVESTOR IN PEOPLE

# CONTENTS

<i>Acknowledgment</i>	vii
Introduction	1
1. Review of the Problem Area	5
2. Adaptive Methods of Fuzzy Clustering	25
3. Kohonen Maps and their Ensembles for Fuzzy Clustering Tasks	51
4. Simulation Results and Solutions for Practical Tasks	79
Conclusion	95
<i>References</i>	97

This page intentionally left blank

# ACKNOWLEDGMENT

The authors would like to thank their families and colleagues for comprehensive support.

Several people put much effort to make this book better.

The team of Emerald Publishing has done a great job of preparing the manuscript. They spent hours of their own time carefully reading through the material, finding typos, and making valuable substantive recommendations.

Each contributor spent many hours helping clarify the goals and concepts of the book, organize the material and generally make the text more intelligible and exciting.

Finally, we should also mention that this book was partially supported by the Research Association of Modern Education and Computer Science (RAMECS) and self-determined research funds of CCNU from the colleges' primary research and operation of MOE (CCNU19TS022).

This page intentionally left blank

# INTRODUCTION

An exponential growth in the amount of stored and processed data has resulted in a rapid transformation of computational intelligence from a disparate set of heuristic methods into an applied discipline that is in high demand and is capable of transforming all human activities. With the use of intelligent systems becoming widespread, the requirement to make them flexible and stable has grown too, which means that they should be steady to any kind of data, adaptive to altering conditions, and must demonstrate transparency for the interpretation of their results. These properties can be administered with the help of arduous mathematical algorithms. The primary attribute of the mathematical methods must be that they should be capable of optimizing the objective functions.

To solve many of the problems that are critical for classification errors, it is not only the quality of the classification that is of importance but also the ability to present the information about the decision-making process in a human-readable form, for example, in the form of a set of rules. In this case, the systems built by a combination of the neural network approach and fuzzy logic (the neuro-fuzzy systems) can significantly expand the range of tasks to be solved. The essential features of fuzzy systems also include the possibility

of obtaining a fuzzy classification at the output, which in practice often avoids making wrong decisions; for example, in cases of uncertainty. Neural networks and neuro-fuzzy systems, as well as adequate learning and self-learning procedures used for solving practical data classification problems, allow to significantly automate this process without requiring an expert in the subject area of the problem to be solved.

Unsupervised learning is one of the most popular areas of computational intelligence because it does not require any preliminary data marking, which can be almost unfeasible while dealing with data volumes typical for up-to-date tasks. One of the primary tasks for unsupervised learning is clustering.

Currently, theoretical and applied research in the field of cluster analysis is characterized by increased attention to the problems of clustering data of arbitrary character under conditions of current and a priori uncertainty that may also contain data with significantly overlapping clusters. For now, many types of intelligent clustering systems (especially hybrid neuro-fuzzy systems) have been developed.

Data mining under conditions of structural and parametric uncertainty is one of the significant problems while dealing with artificial intelligence. The most effective methods for solving this problem have to do with the area of intelligent control, which is based on artificial neural networks (Jang, 1993; Kohonen, 2001; LeCun, Kavukvuoglu, & Farabet, 2010) and fuzzy inference systems (Bezdek, 1981; Kosko, 1992; Zadeh, 1965). In this case, an extensive class of problems such as emulation, identification, prediction, classification, fault detection, and clustering should be solved during the analysis procedure.

The clustering task is especially important because it is solved without supervision. At the same time, the clustering problem is complicated by the fact that the clusters tend to

have a rather complicated shape, substantial overlapping, inside data distributed with different densities, and different volumes for the clusters. Traditional methods for solving the clustering problems (Cover, 1968; Hoepfner, Klawonn, Kruse, & Runkler, 2000; MacQueen, 1965; Ruspini, 1969; Xu & Wunsch, 2009) are based on a priori assumptions about the nature of data distribution and the properties of a sample. Therefore, when it's required to process a growing dataset sequentially (their properties may change over time), it becomes reasonable to use complex hybrid methods based on neural networks (Bodyanskiy, 2005), collective inference methods (Das, Abraham, & Konar, 2009; Dietterich, 2000), and type-2 fuzzy systems (Mendel, 2001; Zadeh, 1975), which are capable of operating with a higher degree of uncertainty for input data.

An analyst should have a vast knowledge of mathematical methods and must have a sound background on fuzzy logic and clustering methods. This is because of the need for a variety of these methods and a high number of parameters to be tuned. Several new effective clustering methods have been developed for solving specific tasks, but they cannot be generalized for other cases (as a matter of fact, it has to do with data processing in a sequential mode and using some advanced approaches to fuzzy logic). Therefore, universal and unified methods for adaptive clustering in an online mode under high-level a priori conditions and the prevailing uncertainty in the features of data distribution should be developed. Neuro-fuzzy systems built on these principles, which are capable of automatic adaptation to the task's conditions, seem promising.

To date, many types of original classification and clustering systems have been developed, including hybrid architectures of neuro-fuzzy systems. It should be noted that the classical neuro and neuro-fuzzy methods have significant

limitations, including low convergence rate (due to the use of learning procedures based on the back-propagation method), the impossibility of functioning under conditions of significant overlapping of data clusters, and the lack of a priori information (for example, an unknown number of clusters). Some existing neural networks do not have any neuro-fuzzy analogs.

Many actual tasks cannot be solved with the help of classical statistical methods of data processing and analysis, which require the development and application of more flexible architecture and methods. Due to the factors mentioned above, the task of developing new hybrid neuro-fuzzy networks, effective methods for their learning and self-learning, which can function in conditions of a priori and current uncertainty regarding the nature of distribution and significant overlapping of data clusters, seems to be quite relevant.

# CHAPTER 1

## REVIEW OF THE PROBLEM AREA

### 1.1. LEARNING AND SELF-LEARNING PROCEDURES

A learning process always involves a system which can be described by a vector of parameters to be learned  $c = (c_1, \dots, c_m)^T$  and input data which are a sequence of observations  $X = \{\vec{x}(k) | 0 \leq k \leq N\}$ , where an  $n$ -dimensional feature vector gives observation for each object  $\vec{x}(k) = (x_1(k), \dots, x_n(k))^T$ . We can define a function describing a system's error by Eq. 1.1 (Tsypkin, 1970):

$$E(c) = \int_X J(x, c) p(x) dx \quad (1.1)$$

where  $J(x, c)$  is some predefined objective function and  $p(x)$  is the density distribution for  $x$  in  $X$ .

For a continuous function  $E(x, c)$ , the purpose of learning procedure is to achieve an optimum state for the system  $c = c^*$  when the functional (Eq. 1.1) attains an extremum value  $\nabla E(c) = \nabla_c J(x, c) = \vec{0}$ . Usually, the state  $c^*$  cannot be determined accurately due to the lack of

information, as, in general, the distribution density  $p(x)$  is also unknown.

If some information about the desired reaction for the system is available for a subset  $X_T \in X$ , it is called a learning set (a training set), and responses  $y(x), x \in X_T$  corresponding to each element are called a learning signal (a training signal). Thus the purpose of learning procedure, in this case, is to minimize the differences between a system's actual output  $\hat{y}(x, c)$  and desired output  $y(x)$ . The objective function  $J(x, c)$  which minimizes the functional (Eq. 1.1) can be defined (in the purest form) as

$$J(x, c) = \|y(x) - \hat{y}(x, c)\|^2. \quad (1.2)$$

Analytical or numerical minimization of these objective functions leads to a variety of supervised learning algorithms.

If the training signal  $y(x)$  is unavailable, construction of the objective function becomes less trivial. In the most general terms, the self-learning process aims to minimize a divergence between the actual density function  $p(x)$  and the approximated one  $\hat{p}(x, c)$  based on the system's performance:

$$J(x, c) = (\hat{p}(x, c) - p(x))^2.$$

Since neither  $p(x)$  nor  $\hat{p}(x, c)$  can be measured directly, it leads to a wide variety of objective functions and algorithms based on these functions.

The main feature of intelligent systems is their ability of learning and self-learning, i.e., the ability to make generalizations based on available and incoming data. This fact allows them to be used for solving problems automatically under

specific conditions, such as lack of a priori information about the data nature and the subject area.

Learning procedures can be described in the form of stochastic difference and differential equations for tunable parameters of a system. In some cases, these equations have an exact solution, but numerical methods are commonly used to ensure an asymptotic convergence to an optimal solution. This leads to the fact that most of the learning procedures are iterative.

Most of the learning procedures can be attributed to either of these two basic classes: supervised learning and unsupervised learning (self-learning). In the supervised learning case, the data contain both input information and examples of desired system responses to the input data that make it possible to train the system by comparing its output signal with samples. In case of unsupervised learning, the system has no information about desired outputs, and its task is to detect patterns in a dataset when any data element is not a solution.

## 1.2. CLUSTERING

Clustering (automatic classification) is one of the primary tasks in data mining, and it implies isolation of similar observations in a dataset in the most general case. In the formal form, the clustering problem is formulated as follows: given a data sample  $X$  consisting  $N$  observations  $X = \{\vec{x}(k) | 0 \leq k \leq N\}$  where each observation is an  $n$ -dimensional feature vector,  $\vec{x}(k) = (x_1(k), \dots, x_i(k), \dots, x_n(k))^T$ . It is often convenient to have a data sample in a matrix form  $X = \{x_{ki}\}$ ,  $1 \leq k \leq N$ ,  $1 \leq i \leq n$ . These forms are similar.

A solution for the clustering problem is to find a partition matrix  $U = \{u(k)_j\}$ ,  $1 \leq k \leq N$ ,  $1 \leq j \leq m$  where  $u(k)_j = 1$

stands for an observation's membership level  $x(k)$  to the  $j$ -th cluster,  $1 \leq j \leq m \ll N$ . A general formulation of the problem does not regulate whether some clusters are set beforehand or found by an algorithm.

The feature that differentiates this problem statement from the classification is that no membership tag is specified for a group for any data subset, i.e., clustering is an unsupervised task.

Decision for clustering problems is fundamentally ambiguous. The various reasons for this are as follows:

- (1) There is no best or universal quality criterion or objective function for a clustering problem. However, there is a vast number of heuristic criteria, as well as some algorithms without a clearly expressed criterion, and all of them can give different results.
- (2) Some clusters are usually previously unknown and are set based on some subjective considerations.
- (3) A clustering result strongly depends on a metric which is typically subjective and determined by an expert.
- (4) Evaluation of the clustering quality is also subjective.

### 1.2.1. Clustering Methods

Although there are a lot of clustering approaches, this book mainly focuses on prototype-based methods (Borgelt, 2005; Xu & Wunsch, 2009). These methods select a small number of the most typical (averaged) observations (also called prototypes or centroids) from a sample (or generating data based on it) and divide the rest of the sample into clusters, based on their proximity.

According to the clustering problem statement, the sample should be divided into clusters, with similar observations