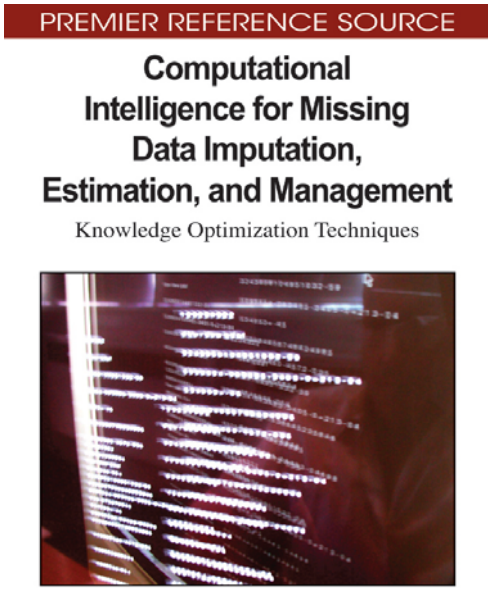
TSHILIDZI MARWALA
COMPUTATIONAL INTELLIGENCE FOR MISSING DATA IMPUTATION, ESTIMATION, AND MANAGEMENT:
KNOWLEDGE OPTIMIZATION TECHNIQUES



Information Science Reference
Hershey, New York
ISBN-10: 1-605-66336-0
ISBN-13: 978-1-60566-336-4
306 pages
2009

Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques addresses a problem largely avoided in engineering mainly because of its difficulty. The book focuses on methods to estimate missing values given to observed data. Providing a defining body of research valuable to those involved in the field of study, this book presents current and new computational intelligence techniques that allow computers to learn the underlying structure of data. The methods covered in this book are:

- the multi-layer perceptron model;
- radial basis functions;
- Gaussian mixture models;
- rough sets;
- support vector machines;
- decision trees;
- fuzzy ARTMAP and extension neural networks.

Many optimization methods are also used:

- scaled conjugate gradient optimization method;
- the expectation maximization algorithm;
- genetic algorithms;
- particle swarm optimization;
- hill climbing;
- simulated annealing.

In *Chapter I*, traditional missing data issues, such as missing data patterns and mechanisms, are described. In *Chapter II*, a method is presented that aimed at approximating missing data and, thereby, capturing variables' interrelationships by combining genetic algorithms and auto-associative neural networks. In *Chapter III*, hybrid auto-associative neural networks models are developed and used in conjunction with genetic algorithms to estimate missing data. In *Chapter IV*, two techniques, i.e., Gaussian mixture models trained using the expectation maximization algorithm and the combined auto-associative neural networks and particle swarm optimization methods are implemented for missing data estimation and then compared. *Chapter V* investigates an imputation technique based on rough sets computation. In *Chapter VI* auto-associative neural networks, principal components analysis and support vector regression are all combined with genetic algorithms, and then used to impute missing variables. In *Chapter VII*, a committee of networks is introduced for missing data estimation. This committee of networks consists of a multi-layer perceptron, support vector machine and radial basis function. Various techniques to handle missing data are discussed in *Chapter VIII*. First, a novel algorithm that classifies and regresses in the presence of missing data is proposed. The algorithm is tested for both classification and regression problems. Second, an estimation algorithm that uses an ensemble of regressors within the context of the boosting mechanism is proposed. Hybrid

genetic algorithms and fast simulated annealing are used to predict missing values and the results are compared. In *Chapter IX*, a classifier method is presented that is based on a missing data estimation framework, and which uses auto-associative multi-layer perceptron neural networks and genetic algorithms. In *Chapter X*, various optimization methods are compared with the aim of optimizing the missing data estimation equation, which is made out of the auto-associative neural networks with missing values as design variables. These optimization techniques are the genetic algorithms, particle swarm optimization, hill climbing and simulated annealing. *Chapter XI* introduces a novel paradigm to impute missing data that combines decision trees with an auto-associative neural networks and principal components analysis. This is designed to answer the crucial question on whether the optimization bounds actually matter in the estimation of missing data. *Chapter XII* presents a control mechanism to access the effect of a demographic variable, *education level*, on the HIV risk of individuals. In *Chapter XIII*, a computational intelligence approach to predicting missing data in the presence of concept drift is presented, using an ensemble of multi-layer perceptron feed-forward neural networks.

Presented book gives a good balance between theory and application of various missing data estimation techniques. Examples from the literature and case studies are used to illustrate the effectiveness of the presented missing data estimation methods.

This book is intended for researchers and practitioners who use data analysis to build decision support systems. In particular the target audience includes engineers, scientists and statisticians. The areas of engineering where decision support tools are becoming widely used are aerospace, mechanical, civil, biomedical and electrical engineering. Furthermore, researchers in statistics and social science will also find the techniques introduced in this book to be highly applicable to their work.

Table of Contents