Preface

Knowledge Discovery and Data Mining (KDD) has become one of the major research areas in data science and data analytics. The more the complexity of the data grows the more grows the need of sophisticated KDD techniques to reveal the knowledge in the data. On the one hand, traditional data analysis tools and techniques have been struggling to tackle this problem, while on the other hand, as a solution biological inspired data mining techniques have been introduced to perform better on different data mining tasks. Data clustering, classification rules mining, association rule mining, sequential pattern mining, outlier detection, feature selection, and recommender systems are some of the area, where biologically inspired techniques have been used. The applications of these techniques include but are not limited to healthcare data, environmental data, sensor data, Web data, semi-structured data, microarray data, and streaming data. Some of the commonly used bio-inspired techniques include Neural Networks, Fuzzy Systems, Genetic Algorithms, Ant Colony Optimization, Particle Swarm Optimization, Artificial Immune Systems, Culture Algorithms, Social Evolution, and Artificial Bee Colony Optimization.

This book is motivated by the desire of bridging the gap between two contemporary fields of Knowledge Discovery and Data Mining (KDD) and biologically inspired optimization techniques. The aim of this book is to highlight the contemporary research in the area of biologically inspired techniques in different data mining domains and the implementation of these techniques in real life data mining problems. The book includes state-of-the-art work in this area and shares the good practices that have enabled this area to grow and flourish. The book provides quality work from established researchers that can be used by the new researchers in the area. Below are highlights of the book.

Chapter 1 introduces some of the main biologically inspired techniques with discussion on advantages and limitations of some of the commonly used techniques. Brief literature overviews of the KDD techniques from Particle Swarm Optimization (PSO) and a Swarm Intelligence (SI)-based optimization are presented. Some of the most cited work significantly impacting work in the area has been overviewed.

Chapter 2 titled, “Probabilistic Control and Swarm Dynamics in Mobile Robots and Ants,” discusses probabilistic control of mobile robots navigating in random environments and mimicking the foraging activity of the ants, a widely accepted optimal method with respect to the environmental conditions. The system was tested using mobile agents based on the Braitenberg vehicles equipped with four types of sensors, mimicking sensing abilities of including short- and long-distance sensing of environmental states, sensing of neighboring agents, and sensing the pheromone traces. The outcomes show the similarity of suggested mobile agents, running both individually and in groups, is statistically indistinguishable from the foraging behavior of real ants observed in laboratory experiments.
Chapter 3, “A Measure Optimized Cost-Sensitive Learning Framework for Imbalanced Data Classification,” implements Particle Swarm Optimization (PSO), a well-known Swarm Intelligence (SI)-based technique, to solve the class imbalanced data classification problem, which is one of the contemporary challenging problems. The authors present an effective wrapper framework incorporating the evaluation measure into the objective function for improving the performance of classification by simultaneously optimizing the best pair of feature subset, intrinsic parameters, and misclassification cost parameter. For evaluation purposes, support vector machine neural networks have been used. Using the standard benchmark classification datasets, the proposed method is compared to commonly used sampling techniques.

Chapter 4 of the book, “Towards an Improved Ensemble Learning Model of Artificial Neural Networks: Lessons Learned on Using Randomized Numbers of Hidden Neurons,” describes an application area of Artificial Neural Networks (ANN) using ensemble learning model. The authors argue that ensemble machine learning is capable of improving performance of instability of ANN. The study presents using a linear search and randomized assignment of the number of hidden neurons as an extension to the already presented approaches in the area. Using standard model evaluation criteria and novel ensemble combination rules, the results of this study suggest that having a large number of “unbiased” randomized guesses of the number of hidden neurons beyond 50 performs better than very few occurrences of those that were optimally determined.

Chapter 5, “Ant Programming Algorithms for Classification,” presents a detailed literature overview of ant programming approaches to extract comprehensible classifiers. There exist three such algorithms for classification rule mining: two of them for regular classification and one for imbalanced classification. These algorithms mainly differ in the optimization approach, single-objective or multi-objective, and the purpose of the classification task. The chapter collects these algorithms, presenting different experimental studies that confirm the aptitude of this metaheuristic to address the classification task.

Chapter 6, “Machine Fault Diagnosis and Prognosis using Self-Organizing Map,” proposes a Self-Organizing Map (SOM) method for fault diagnosis and prognosis of manufacturing systems, machines, components, and processes. The study concentrates on optimizing the discovery process of machine health. The proposed method provides a way to predict manufacturing faults, which can be used to develop a good maintenance strategy for the optimum number of corrective maintenance. Self-organizing maps have been shown useful to classify the fault and predict the degrading of machines, components, and processes effectively, clearly, and easily.

Chapter 7, “An Enhanced Artificial Bee Colony Optimizer for Predictive Analysis of Heating Oil Prices using Least Squares Support Vector Machines,” uses one of the emerging technique from Swarm Intelligence, called Artificial Bee Colony (ABC) optimization, for prediction analysis. The proposed approach is based on the hybridization of Least Squares Support Vector Machines (LSSVM) with an enhanced Artificial Bee Colony (eABC) technique. An interesting application area of the research “oil prices prediction” has been proposed in this study. The technique works in two phases. Initially, a Levy mutation is introduced to keep the model from falling into local minimum; later, the predictive analysis is succeeded by the LSSVM. Realized in predictive analysis of heating oil prices, the empirical findings manifest the superiority of eABC-LSSVM in prediction accuracy.
Chapter 8, “Comparison of Linguistic Summaries and Fuzzy Functional Dependencies Related to Data Mining,” overviews the task of performing data mining based on fuzzy logic. The chapter explores the ways to discover potentially useful knowledge from relational databases by integrating fuzzy functional dependencies and linguistic summaries. The finding of this work says that using fuzzy functional dependencies and linguistic summaries in a complementing way could mine a variety of valuable information from relational databases, which can result in decision-making, reducing the number of attributes in databases, and estimating missing values.

Chapter 9 explores the research work in the area of genetic programming for civil engineering. The chapter, “Application of Artificial Neural Network and Genetic Programming in Civil Engineering,” investigates Genetic Programming (GP) and Artificial Neural Network (ANN) models for prediction of air entrainment rate (QA) of triangular sharp-crested weir. The authors present a comparative study between the developed Back Propagation (BP), Generalized Regression Neural Network (GRNN), and GP models.

Chapter 10, “A Promising Direction towards Automatic Construction of Relevance Measures,” quantifies a given feature for its relevancy to a target variable. Measuring relevance and redundancy is a central concept in feature selection and features reduction. The authors identify the lack of generality as a core problem and show that regardless of the type of heuristic measure and search strategy, heuristic methods cannot optimize the performance of all learning algorithms. They propose a hyper-heuristic method that through an evolutionary process automatically generates an appropriate relevance measure for a given problem. The new approach can detect relevant features in difficult scenarios.

In chapter 11, Earliest Deadline First (EDF) and Ant Colony Optimization (ACO) have been used for adaptive scheduling of the distributed systems. The chapter titled, “Adaptive Scheduling for Real-Time Distributed Systems,” proposes the use of Artificial Intelligence methods towards more realistic domains requiring real-time responses. The finding of the research reports that Ant Colony Optimization (ACO) performs quite well for scheduling real-time distributed systems during overloaded conditions.

Chapter 12, “Discovery of Emergent Sorting Behavior using Swarm Intelligence and Grid-Enabled Genetic Algorithms,” uses simple local comparison and swap operators for sorting sequences. The authors present the basic concepts and an experimental validation of Emerge-Sort, a sorting algorithm that does not depend on being told which way a sequence should be sorted and yet manages to sort that sequence based on randomly applied simple local operators. The authors experimentally validate square run-time behavior for emergent sorting and suggest that not knowing in advance which direction to sort and allowing such direction to emerge imposes a penalty over conventional techniques.

Chapter 13, “Application of Biologically Inspired Techniques for Industrial and Environmental Research via Air Quality Monitoring Network,” proposes a systematic methodology for simultaneous identification of an emission source and emission rate. The authors assess the application of artificial neural networks for this purpose.

Chapter 14, “Online Prediction of Blood Glucose Levels using Genetic Algorithm,” presents an interesting application area of the genetic algorithms. The authors propose a computer-based system that predicts Blood Glucose Level (BGL) in diabetic patients so that they can manage their blood glucose level.
The last chapter of the book, “Security of Wireless Devices using Biological-Inspired RF Fingerprinting Technique,” proposes the use of ANN for a growing area of communication security. Radio Frequency (RF) fingerprinting, a means of providing an additional layer of security for wireless devices, is a security mechanism inspired from biological fingerprint identification systems. RF fingerprinting classification is performed by selecting an “unknown” signal from the pool, generating its RF fingerprint, and using a classifier to correlate the received RF fingerprint with each profile RF fingerprint stored in the database. The authors evaluate the performance of the KNN and neural network classification in three different scenarios. The experimental study compares the outcome for these different scenarios.

Shafiq Alam  
University of Auckland, New Zealand

Gillian Dobbie  
University of Auckland, New Zealand

Yun Sing Koh  
University of Auckland, New Zealand

Saeed ur Rehman  
Unitec Institute of Technology, New Zealand