Preface

Swarm Intelligence (SI) is an innovative distributed intelligent paradigm for solving optimization problems that originally took its inspiration from the biological examples by swarming, flocking and herding phenomena in vertebrates. Particle Swarm Optimization (PSO) incorporates swarming behaviors observed in flocks of birds, schools of fish, or swarms of bees, and even human social behavior, from which the idea is emerged. Ant Colony Optimization (ACO) deals with artificial systems that is inspired from the foraging behavior of real ants, which are used to solve discrete optimization problems.

Historically the notion of finding useful patterns in data has been given a variety of names including data mining, knowledge discovery, information extraction, etc. Data Mining is an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. In order to achieve this, data mining uses computational techniques from statistics, machine learning and pattern recognition.

Data mining and *Swarm intelligence* may seem that they do not have many properties in common. However, recent studies suggests that they can be used together for several real world data mining problems especially when other methods would be too expensive or difficult to implement.

This book deals with the application of swarm intelligence methodologies in data mining. Addressing the various issues of swarm intelligence and data mining using different intelligent approaches is the novelty of this edited volume. This volume comprises of 11 chapters including an introductory chapters giving the fundamental definitions and some important research challenges. Chapters were selected on the basis of fundamental ideas/concepts rather than the thoroughness of techniques deployed. The eleven chapters are organized as follows.

In Chapter 1, *Grosan et al.* present the biological motivation and some of the theoretical concepts of swarm intelligence with an emphasis on particle swarm optimization and ant colony optimization algorithms. The basic data mining terminologies are explained and linked with some of the past and ongoing works using swarm intelligence techniques.

Martens et al. in Chapter 2 introduce a new algorithm for classification, named AntMiner+, based on an artificial ant system with inherent self-organizing capabilities. AntMiner+ differs from the previously proposed AntMiner classification technique in three aspects. Firstly, AntMiner+ uses a MAX-MIN ant system which is an improved version of the originally proposed ant system, yielding better performing classifiers. Secondly, the complexity of the environment in which the ants operate has substantially decreased. Finally, AntMiner+ leads to fewer and better performing rules.

In Chapter 3, *Jensen* presents a feature selection mechanism based on ant colony optimization algorithm to determine a minimal feature subset from a problem domain while retaining a suitably high accuracy in representing the original features. The proposed method is applied to two very different challenging tasks, namely web classification and complex systems monitoring.

Galea and *Shen* in the fourth chapter present an ant colony optimization approach for the induction of fuzzy rules. Several ant colony optimization algorithms are run simultaneously, with each focusing on finding descriptive rules for a specific class. The final outcome is a fuzzy rulebase that has been evolved so that individual rules complement each other during the classification process.

In the fifth chapter *Tsang and Kwong* present an ant colony based clustering model for intrusion detection. The proposed model improves existing ant-based clustering algorithms by incorporating some meta-heuristic principles. To further improve the clustering solution and alleviate the curse of dimensionality in network connection data, four unsupervised feature extraction algorithms are also studied and evaluated.

Omran et al. in the sixth chapter present particle swarm optimization algorithms for pattern recognition and image processing problems. First a clustering method that is based on PSO is discussed. The application of the proposed clustering algorithm to the problem of unsupervised classification and segmentation of images is investigated. Then PSO-based approaches that tackle the color image quantization and spectral unmixing problems are discussed.

In the seventh chapter *Azzag et al.* present a new model for data clustering, which is inspired from the self-assembly behavior of real ants. Real ants can build complex structures by connecting themselves to each others. It is shown is this paper that this behavior can be used to build a hierarchical tree-structured partitioning of the data according to the similarities between those data. Authors have also introduced an incremental version of the artificial ants algorithm.

Kazemian et al. in the eighth chapter presents a new swarm data clustering method based on Flowers Pollination by Artificial Bees (FPAB). FPAB does not require any parameter settings and any initial information such as the number of classes and the number of partitions on input data. Initially, in FPAB, bees move the pollens and pollinate them. Each pollen will grow in proportion to its garden flowers. Better growing will occur in better conditions. After some iterations, natural selection reduces the pollens and flowers and the gardens of the same type of flowers will be formed. The prototypes of each gardens are taken as the initial cluster centers for Fuzzy C Means algorithm which is used to reduce obvious misclassification errors. In the next stage, the prototypes of gardens are assumed as a single flower and FPAB is applied to them again.

Palotai et al. in the ninth chapter propose an Alife architecture for news foraging. News foragers in the Internet were evolved by a simple internal selective algorithm: selection concerned the memory components, being finite in size and containing the list of most promising supplies. Foragers received reward for locating not yet found news and crawled by using value estimation. Foragers were allowed to multiply if they passed a given productivity threshold. A particular property of this community is that there is no direct interaction (here, communication) amongst foragers that allowed us to study compartmentalization, assumed to be important for scalability, in a very clear form.

Veenhuis and Köppen in the tenth chapter introduce a data clustering algorithm based on species clustering. It combines methods of particle swarm optimization and flock algorithms. A given set of data is interpreted as a multi-species swarm which wants to separate into single-species swarms, i.e., clusters. The data to be clustered are assigned to datoids which form a swarm on a two-dimensional plane. A datoid can be imagined as a bird carrying a piece of data on its back. While swarming, this swarm divides into sub-swarms moving over the plane and consisting of datoids carrying similar data. After swarming, these sub swarms of datoids can be grouped together as clusters.

In the last chapter *Yang et al.* present a clustering ensemble model using ant colony algorithm with validity index and ART neural network. Clusterings are visually formed on the plane by ants walking, picking up or dropping down projected data objects with different probabilities. Adaptive Resonance Theory (ART) is employed to combine the clusterings produced by ant colonies with different moving speeds.

We are very much grateful to the authors of this volume and to the reviewers for their tremendous service by critically reviewing the chapters. The editors would like to thank Dr. Thomas Ditzinger (Springer Engineering Inhouse Editor, Studies in Computational Intelligence Series), Professor Janusz Kacprzyk (Editor-in-Chief, Springer Studies in Computational Intelligence Series) and Ms. Heather King (Editorial Assistant, Springer Verlag, Heidelberg) for the editorial assistance and excellent cooperative collaboration to produce this important scientific work. We hope that the reader will share our excitement to present this volume on '*Swarm Intelligence in Data Mining*' and will find it useful.

April, 2006

Ajith Abraham, Chung-Ang University, Seoul, Korea
Crina Grosan, Cluj-Napoca, Babeş-Bolyai University, Romania
Vitorino Ramos, Technical University of Lisbon, Portugal