Special Issue on Computational Intelligence in Data mining

In our society the amount of data doubles almost every year. Hence, there is an urgent need for a new generation of computationally intelligent techniques and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volume of data.

When we attempt to solve real-world problems, like extracting knowledge from large amount of data, we realize that they are typically ill-defined systems, difficult to model and with large-scale solution spaces. In these cases, precise models are impractical, too expensive, or non-existent. Furthermore, the relevant available information is usually in the form of empirical prior knowledge and input–output data representing instances of the system's behavior. Therefore, we need an approximate reasoning system capable of handling such imperfect information. While Bezdek [2] defines such approaches within a frame called computational intelligence, Zadeh [3] explains the same using the soft computing paradigm. According to Zadeh "... in contrast to traditional, hard computing, soft computing is tolerant of imprecision, uncertainty, and partial truth." In this context Fuzzy Logic (FL), Probabilistic Reasoning (PR), Neural Networks (NNs), and Evolutionary Algorithms (EAs) are considered as main components of CI. Each of these technologies provide us with complementary **reasoning** and **searching** methods to solve complex, real-world problems. What is important to note is that soft computing is not a melange. Rather, it is a partnership in which each of the partners contributes a distinct methodology for addressing problems in its domain. In this perspective, the principal constituent methodologies in CI are complementary rather than competitive [4].

This special issue deals with the importance of computational intelligence (CI) paradigms in data mining and knowledge discovery.

The first paper is aimed to give a comprehensive view about the links between computational intelligence and data mining. Further, a case study is also given in which the extracted knowledge is represented by fuzzy rule-based expert systems obtained by soft computing based data mining algorithms. It is recognized that both model performance and interpretability are of major importance, and effort is required to keep the resulting rule bases small and comprehensible. Therefore, CI technique based data mining algorithms have been developed for feature selection, feature extraction, model optimization and model reduction (rule base simplification). The results illustrate that that CI based tools can be applied in a synergistic manner though the nine steps of knowledge discovery.

The remaining papers were selected from the papers presented at the **4th International Conference on Intelligent Systems Design and Application (ISDA'04 at August 26-28, 2004)** on the basis of fundamental ideas/concepts rather than the thoroughness of techniques deployed.

The second paper of this special issue by **Górriz, Se-gura, Puntonet and Salmerón** presents a survey of forecasting preprocessing techniques. Authors have illustrated that these methods could play a major role in the final model accuracy.

Many computational intelligence method can be used for clustering purposes beside the classical techniques. **Liu, Özyer, Alhajj and Barker** in the third paper propose a new clustering algorithm integrating multi-objective genetic algorithm and validity analysis.

In the fourth paper, **Podgorelec, Kokol, Heričko and Rozman** present a method for data classification which is based on constructing decision graphs with the help of several agents. They present a two-leveled evolutionary algorithm for the induction of decision graphs and describe the principle of classification based on the decision graphs.

Kotsiantis and Pintelas in the fifth paper combine a simple Bayesian classification method with Logitboost, which is a bias reduction technique. Logitboost requires a regression algorithm for base learner. For this reason, they slightly modify simple Bayesian classifier in order to be able to run as a regression method. They performed a large-scale comparison with other state-of-the-art algorithms and ensembles on 27 standard benchmark datasets and the empirical results looks very interesting.

In the sixth paper, **Wang and Garibaldi** propose a method useful in cancer diagnosis called Simulated Annealing Fuzzy Clustering. This technique can solve two major problems that exist in simpler methods: it does not need the number of clusters in advance and is able to avoid sub-optimal solutions in some sense.

Ivancsy and Vajk in the seventh paper focus on mining frequent patterns in large transactional databases. The contribution of their new method is to count the short patterns in a very fast way, using a specific index structure.

An important research field within pattern recognition is the analysis of time series data. In the eighth paper, **Toshniwal and Joshi** introduce a novel approach for performing similarity search in time series data. Their technique is based on the intuition that similar time sequences will have similar variations in their slopes.

Chong, Abraham and Paprzycki in the ninths paper propose a hybrid model involving decision trees and neural networks for classification of the type of injury severity of various traffic accidents. They also consider neural networks trained using hybrid learning approaches, support vector machines, and decision trees.

Ahvenlampi and Kortela in the last paper propose a hybrid system for controllability of quality in continuous digesters in which Self-Organizing Maps and Gustafson-Kessel fuzzy clustering algorithm are used.

The editors wish to thank Professor Matjaz Gams(Editor-in-Chief of Informatica)for providing the opportunity to edit this special issue on Computational Intelligence in Data Mining. We would also like to thank

the referees who have critically evaluated the papers within the short stipulated time. Finally we hope the reader will share our joy and find this special issue very useful.