Raport Badawczy
Research Report

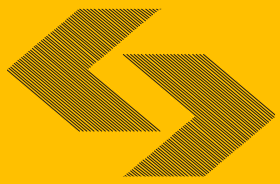
RB/74/2005

Book review: Evolutionary computation in data mining by Ashish Ghosh, Lakhmi C. Jain

J. Stańczak

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences



Control and Cybernetics

vol. **34** (2005) No. 4

Book review:

Evolutionary Computation in Data Mining

by

Ashish Ghosh, Lakhmi C. Jain

"*Evolutionary Computation in Data Mining*" edited by Ashish Ghosh and Lakhmi C. Jain, published in the series on Studies in Fuzziness and Soft Computing by Springer contains twelve skillful papers, which show the advances and state of the art in the domain of evolutionary methods' application in data mining.

Data mining is a relatively broad domain of automatic knowledge discovery from databases and is now one of the most developed fields in the area of artificial intelligence. A rapid growth of data collected in various realms of human activity and their potential usefulness in these realms requires appropriate and efficient tools to extract and use of the potentially gathered knowledge. Evolutionary algorithms are very powerful tools and can be utilized in many ways to make use of knowledge hidden in terabites of data collected in databases all over the world.

In the Chapter 1 by Ashish Ghosh some essential aspects of application of evolutionary computation methods for rule generation are discussed.

Chapter 2 by Jose Ramon Cano, Francisco Herrera and Manuel Lozano focuses on data preprocessing methods meant to minimize the size of data to be processed. Data reduction is obtained using a CHC evolutionary algorithm with stratification strategy to select representative instances of the processed data set. The presented evolutionary method outperforms the non-evolutionary ones and gives better instance reduction rates, higher accuracy and lower resources consumption.

The next chapter by Matthew G. Smith and Larry Bull presents also a preprocessing method, where genetic programming is used to feature construction (inferring new features enables to find hidden relationships between them) and a genetic algorithm for feature selection from the set produced by the first stage. The preprocessing method with the C4.5 classification algorithm works significantly better than the single C4.5 method.

Also the fourth chapter by Riyaz Sikora presents a multi-agent based method of feature selection and inductive learning. Genetic algorithms are treated as learning agents to discover solutions of small subproblems into which the whole data set is decomposed.

Next three chapters show different aspects of evolutionary algorithms' utilization to rule generation. Xiju Fu and Lipo Wang in Chapter 5 introduce a neuro-evolutionary system which extracts rules from trained RBF neural network with a gradient descent method, using genetic algorithm to determine the feature subsets of different classes. Chapter 6 by Qi Yu, Kay Chen Tan and Tong Heng Lee describes a coevolutive method - CORE (Coevolutionary Rule Extractor), where populations of rules and rule sets coevolve concurrently. The results obtained show that this method generates cohesive and comprehensive rule sets. In the next chapter Minh Ha Nguyen, Hussein Abbas and Robert McKay use an ensemble of different neural networks to improve the generalization ability of learning machines, which is crucial for rule discovery. Different methods of creating an ensemble are discussed and proposed: - simple evolutionary computation with no ensemble; - memetic, through back propagation, evolutionary computation with no ensemble; - simple evolutionary computation with ensemble; - ensemble built using an island model; - ensemble built using evolutionary computation with negative correlation learning; - ensembles built using evolutionary Pareto multi-objective optimization.

In Chapter 8, Olfa Nasraoui, Elizabeth Leon and Raghu Krishnapuram introduce a Unsupervised Niche Clustering algorithm (UNC). It is a very good tool for unsupervised clustering and is able to automatically determine the number of clusters. The program is based on an idea of multimodal function optimization using genetic niching methods.

Chapter 9 by Ajith Abraham deals with evolutionary computation utilized for intrusion detection and also to discover useful information from the secondary data obtained from the interactions of the users with the web. In the system presented evolutionary computation is used to optimize discovered data clusters for a fuzzy inference system.

Next chapter, by W. B. Langdon and S. J. Barrett, describes a genetic programming based method to obtain a predictive model of human bioavailability for testing and discovery *in silico* of new chemicals and drugs. This model can help to decide if some chemicals are worth to be synthesized.

In Chapter 11 Gary B. Fogel presents evolutionary computation methods applied to data mining in the area of DNA microarray analysis, where a large number of genes and their expressions in different cell types provides an immense space of possible relations. Evolutionary methods can satisfactorily cope with this problem.

Chapter 12 by Po-Chang Ko and Ping-Chen Lin deals with some aspects of application evolutionary computations in economy, where forecasting of corporate financial distress plays an increasingly important role. The authors present a modularized financial distress forecasting mechanism based on evolutionary algorithm to extract essential financial patterns.

The book contains a very good survey of evolutionary methods applied in data mining. It is an unique position in the literature and should be very helpful for both specialists and novices interested in this area.

Jaroslaw Stańczak

Ashish Ghosh and Lakhmi C. Jain, eds. *Evolutionary Computation in Data Mining.* Series: Studies in Fuzziness and Soft Computing, vol. 163. Springer-Verlag, 265+XVII pages, 2005. ISBN 3-540-22370-3. Price (hardcover): 89.95 EUR.