Data Mining and Soft Computing

Francisco Herrera
Research Group on Soft Computing and Information Intelligent Systems (SCI²S)
Dept. of Computer Science and A.I.
University of Granada, Spain

Email: herrera@decsai.ugr.es
http://sci2s.ugr.es
http://decsai.ugr.es/~herrera
In this course:

We will introduce the Data Mining and Knowledge Discovery area: steps, task, challenges ..

We will introduce Soft Computing techniques: Fuzzy Logic, Genetic Algorithms, …

… and we will present the use of Soft Computing techniques in Data Mining
Data Mining and Soft Computing

Material of this course at:
http://sci2s.ugr.es/docencia/asignatura.php?id_asignatura=14
Data Mining and Soft Computing

- **Data mining:**
  - Extraction of interesting *(non-trivial, implicit, previously unknown and potentially useful)* information or patterns from data in large databases.

Knowledge

Data Mining

Interpretation

Evaluation

Preprocessing & cleaning

Selection

Target data

Processed data

Patterns
Data Mining

How can I analyze this data?

We have rich data, but poor information

Data mining—searching for knowledge (interesting patterns) in your data.

J. Han, M. Kamber. Data Mining. Concepts and Techniques
Morgan Kaufmann, 2006 (Second Edition)
SoftComputing

Soft computing refers to a collection of computational techniques in computer science, machine learning and some engineering disciplines, which study, model, and analyze very complex phenomena: those for which more conventional methods have not yielded low cost, analytic, and complete solutions.

Prof. Zadeh:

"...in contrast to traditional hard computing, soft computing exploits the tolerance for imprecision, uncertainty, and partial truth to achieve tractability, robustness, low solution-cost, and better rapport with reality”

Lotfi A. Zadeh
Introduce “Fuzzy Logic” in 1965
The Field of Interest of the Society shall be the theory, design, application, and development of biologically and linguistically motivated computational paradigms emphasizing neural networks, connectionist systems, genetic algorithms, evolutionary programming, fuzzy systems, and hybrid intelligent systems in which these paradigms are contained.
Data Mining and Soft Computing

Contents:

Part I. Principles of Data Mining

- Introduction to Data Mining and Knowledge Discovery
- Data Preparation
- Introduction to Prediction, Classification, Clustering and Association
- Data Mining - From the Top 10 Algorithms to the New Challenges
Contents:

Part II. Soft Computing Techniques in Data Mining

- Introduction to Soft Computing. Focusing our attention in Fuzzy Logic and Evolutionary Computation
- Soft Computing Techniques in Data Mining: Fuzzy Data Mining and Knowledge Extraction based on Evolutionary Learning
Data Mining and Soft Computing

Contents:

Part III. Data Mining: Some Advanced Topics

- Some Advanced Topics I: Classification with Imbalanced Data Sets
- Some Advanced Topics II: Subgroup Discovery
- Some advanced Topics III: Data Complexity
- Final talk: How must I Do my Experimental Study? Design of Experiments in Data Mining/Computational Intelligence. Using Non-parametric Tests. Some Cases of Study.
Bibliography

J. Han, M. Kamber.
Data Mining. Concepts and Techniques
Morgan Kaufmann, 2006 (Second Edition)
http://www.cs.sfu.ca/~han/dmbook

I.H. Witten, E. Frank.
Data Mining: Practical Machine Learning Tools and Techniques,

Pang-Ning Tan, Michael Steinbach, and Vipin Kumar
Introduction to Data Mining (First Edition)
Addison Wesley, (May 2, 2005)

Dorian Pyle
Data Preparation for Data Mining
Morgan Kaufmann, Mar 15, 1999

Mamdouh Refaat
Data Preparation for Data Mining Using SAS
Morgan Kaufmann, Sep. 29, 2006)
Summary

1. Introduction to Data Mining and Knowledge Discovery
2. Data Preparation
3. Introduction to Prediction, Classification, Clustering and Association
4. Data Mining - From the Top 10 Algorithms to the New Challenges
5. Introduction to Soft Computing. Focusing our attention in Fuzzy Logic and Evolutionary Computation
6. Soft Computing Techniques in Data Mining: Fuzzy Data Mining and Knowledge Extraction based on Evolutionary Learning
8. Some Advanced Topics I: Classification with Imbalanced Data Sets
9. Some Advanced Topics II: Subgroup Discovery
10. Some advanced Topics III: Data Complexity