Integral Support to Environmental Decision-Making through GESCONDA

Miquel Sànchez-Marrè[a,b], Karina Gibert[a,c], Beatriz Sevilla[a,b]

[a]*Knowledge Engineering and Machine Learning Group (KEMLG)*
[b]*Computer Software Dept.*
[c]*Statistics and Operations Research Dept.*
Universitat Politècnica de Catalunya-BarcelonaTech, Barcelona, Catalonia
(miquel@lsi.upc.edu, karina.gibert@upc.edu, bea.sevilla@gmail.com)

Abstract: In this work, the GESCONDA system is presented. Initially it was conceived as a system for knowledge discovery and Data Mining, but currently, the system supports two new functionalities. A case-based reasoning engine and a rule-based reasoning shell are provided. These new skills of GESCONDA makes it a suitable prototype tool for the deployment of Intelligent Decision Support Systems, including all main steps like data preparation and filtering, data mining, model validation, reasoning abilities to generate solutions, and predictive models to support final users. The purpose of the paper is to present its architecture as well as its functionalities.

Keywords: Intelligent Decision Support Tools; Intelligent Data Analysis; Environmental Systems

1. INTRODUCTION

GESCONDA is a system for Knowledge Discovery and Data Mining. The aim of this research is to design and develop a prototype tool for intelligent data analysis and implicit knowledge management of databases that can guide the decision-making process, with special focus on databases with high amount and heterogeneous data, such as environmental. Though there are several Data Mining systems, the more relevant of this proposal are the incorporation of the statistical data filtering and the pre-processing in the same software tool together with the intelligent data analysis techniques as well as the interaction of different data mining methods. Either statistical techniques or Artificial Intelligence techniques or even mixed techniques that can cooperate among them are combined and used to extract the knowledge contained within data.

GESCONDA is a standard Java application with a friendly graphical user interface (GUI).

GESCONDA has a four-layer architecture. These four levels are briefly described in the following paragraphs.

Currently, GESCONDA is composed at the data filtering layer by several statistical data filtering analysis methods, such as one-way and two-way descriptive statistics, missing data analysis, relationship between variables, hybrid Artificial Intelligence and Statistical methods.

At the recommendation and meta-knowledge layer there is the possibility of introducing meta-knowledge about variables and examples, some feature weighting techniques are available, and a data mining techniques recommender is implemented to suggest the most suitable DM methods to be used according the aims of the user and the given domain.

In addition, several machine learning techniques, coming from Artificial Intelligence, such as conceptual clustering methods, decision tree induction and classification rule induction at the knowledge discovery layer. Some of them developed by the authors, as detailed in Gibert, (1998), Comas et al. (2001), Sànchez-Marrè et al.(1999), and Gibert (2004).

Finally, in the knowledge management layer, two reasoning engines have been developed: a case-based reasoning engine and a rule-based reasoning shell, where all the knowledge obtained can be combined to solve the fault diagnosis, the repairing actions or the predictive support within the environmental problems.

The project purpose is to extend these intelligent system with some new computational modules, such as soft computing methods, support vector machines approaches, dynamical analysis techniques, and hybrid methods integrating Artificial Intelligence approaches and Statistical ones

The prototype is evolving from a simple data mining and knowledge discovery tool to a more complex intelligent environmental decision support tool, with a high emphasis on environmental features like:

- Huge amount of data
- Incomplete information: many missing values
- Many descriptive features: feature relevance problem
- Temporal / Spatial feature: Dynamic and Spatial data analysis
- Different Data format: Spatial data formats

As an example of the use of GESCONDA, an IEDSS was built-up to supervise a wastewater treatment plant operation as described in Rodríguez-Roda et al. [2002].

REFERENCES

Comas J., Dzeroski S. Gibert, K., Rodríguez-Roda I. and Sànchez-Marrè M. Knowledge Discovery by means of inductive methods in wastewater treatment plant data. AI Communications 14(1):45-62, January 2001.

Gibert K., Cortés U, Aluja, T., Knowledge Discovery with Clustering Based on Rules. Interpreting Results,In Principles of Data Mining and Knowledge Discovery, J. M. Zytkow, M. Quafafou Eds., LNAI 510, p 83-92, Springer-Verlag, 1998.

Gibert K., Flores X. and Sànchez-Marrè M. Comparison of classifications in environmental databases using GESCONDA. In 4th ECAI Workshop on Binding Environmental Sciences and Artificial Inte-lligence (BESAI'2004), 13-1:13-10, at ECAI'2004. València, 2004.

Rodríguez-Roda I., Comas J., Colprim J., Poch M., Sànchez-Marrè M., Cortés U., Baeza, J. and Lafuente J. A hybrid supervisory system to support wastewater treatment plant operation: implementation and validation. Water Science & Technology, 45(4-5), 289-297 (2002).

Sànchez-Marrè M., Gibert K., Rodríguez-Roda R., Bueno E., Mozo L., Clavell A., Martín M. and Rougé P. (2002). Development of an Intelligent Data Analysis System for Knowledge Management in Environmental Data Bases. In Rizzoli, A. E. and Jakeman, A. E. (eds.) Integrated Assessment and Decission Supoort. Proceedings of the First biennial Meeting of the International Environmental Modelling and Software Society, Vol. 3, pp:420-425. iEMSs 2002.

Sànchez-Marrè M., U. Cortés, I. R.-Roda, M. Poch. Sustainable Case Learning for continuous domains. Environmental Modelling & Software 14:349-357, 1999.

Sànchez-Marrè M., Béjar J., Cortés U., Gràcia J., Lafuente J. and Poch M. Concept formation in WWTP by means of classification techniques: a compared study. Applied Intelligence, 7(2), 147-166. 1997.