International Environmental Modelling and Software Society (iEMSs)
2010 International Congress on Environmental Modelling and Software
Modelling for Environment's Sake, Fifth Biennial Meeting, Ottawa, Canada
David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, T. Filatova (Eds.)
http://www.iemss.org/iemss2010/index.php?n=Main.Proceedings

Evolving GESCONDA to an Intelligent Decision Support Tool

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-Barcelona**Tech**, 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

The main goal of this research is to design and develop a tool, named GESCONDA, for intelligent data analysis and management of implicit knowledge from databases and also for providing the users with reasoning capabilities, with special focus on environmental databases and environmental modelling. The latter is remarkable due to the high amount and high heterogeneous data in the environmental field. It will provide support to knowledge discovery and data mining to obtain valid new knowledge as well as to solving tasks and predictive tasks using the obtained knowledge/models to guide the decision-making process.

Although in the literature other knowledge discovery tools or commercial systems exist, but none of them strongly integrates statistical and machine learning methods together nor integrates the problem solving and predictive skills in the same tool. Besides the possibility of explicit management of the produced knowledge in Knowledge Bases or Case Bases, the tool also shows other abilities. There are mixed techniques that can cooperate among them to discover and extract the knowledge contained in data allowing interaction among all methods, and the existence of a recommender agent, which will suggest the best method to be used depending on the target domain and on the goals specified by users. The GESCONDA tool has been successfully used in the partial development of IEDSS in several environmental domains (see Sànchez-Marrè and Gibert [2008] or Sànchez-Marrè et al. [2004]).

1.1 Related Work

Intelligent artificial applications that support decision-making processes and problem-solving activities have proliferated and evolved over the past decades. Nowadays, there is a wide range of applications both from the commercial and academic source. One of the most famous commercial systems is SPSS (www.spss.com) that contains a data mining module previously called Clementine. This software expects to be seen as an intelligent business tool and offers a friendly interface to manage the data and to model some algorithm, such as decision trees, artificial neural nets, support vector machines, regressions and also it allows combine models. There are non-commercial suites. Probably the most popular are Weka and RapidMiner, although RapidMiner offers a free restricted suite and an Enterprise editions that contains own algorithms.

Weka (www.cs.waikato.ac.nz/ml/weka) is an application that supports data pre-processing, classification, regression, clustering, association rules and visualization. Also, Weka is available as an API and other applications have included, such as RapidMiner (rapid-i.com) and Penthalo (www.pentaho.com). RapidMiner is a suite that offers data Integration, analytical ETL, data analysis and reporting.

Keel (www.keel.es) is a software tool to assess evolutionary algorithms for Data Mining problems including regression, classification, clustering, pattern mining, etc. It expects to be a tool for researching or with educational goal that are the reasons why it has a lack in post processing methods how occurs in RapidMiner and Weka.

MSMiner [Shi *et al.*, 2007] is a generic multi-strategy data mining platform for decision support. The goal of research and development is to implement an integrated, extensible decision support tool by employing data warehousing and data mining technologies. MSMiner also consists of four parts: ETL (data extraction, data transformation, data loading) subsystem, metadata management subsystem, data warehouse manager subsystem and data mining subsystem.

Up to now, generic tools have been introduced but any of them it is prepared to support a post process that allows the system to reason and to solve problems. Neither, none of them could deal with some real problems which have more than one attribute to predict or solve.

2. INTELLIGENT DECISION SUPPORT SYSTEMS

An Intelligent Decision Support System (IDSS) can be defined as an intelligent information system for decreasing the decision-making time and improving consistency and quality of decisions as stated by Haagsma & Johanns [1994]. An IEDSS is an ideal decision-oriented tool for suggesting recommendations in an environmental domain. The main outstanding feature of IEDSS is the knowledge embodied, which provides the system with enhanced abilities to reason about the environmental system in a more reliable way.

The high quantity of information and implicit knowledge patterns contained in large database coming from any dynamical environmental process is remarkable. Management of this data is a very difficult task, due to the complexity of the features involved in those systems, such as biological, chemical, physical, electrical, or ecological ones.

Intelligent Environmental Decision Support Systems (IEDSS) integrate the expert knowledge stored by human experts through years of experience in the process operation and management. In addition, some knowledge can be obtained through the intelligent analysis of large databases coming from historical operation of the environmental process. Thus, knowledge mining and knowledge acquisition, as well as reasoning over the acquired models are a key step to build reliable IEDSS.

In this paper, we consider approaches and methods of searching solutions based on structural analogy and cases or in inference rules, which are oriented to use them in real-time (RT-IDSS).

3. GESCONDA SOFTWARE ARCHITECTURE

GESCONDA software was designed (see figure 1) as a 4-layer architecture connecting the user with the environmental system. These levels are:

- Data Filtering:
  - Data cleaning;
  - Missing data analysis and management;
  - Outlier data analysis and management;

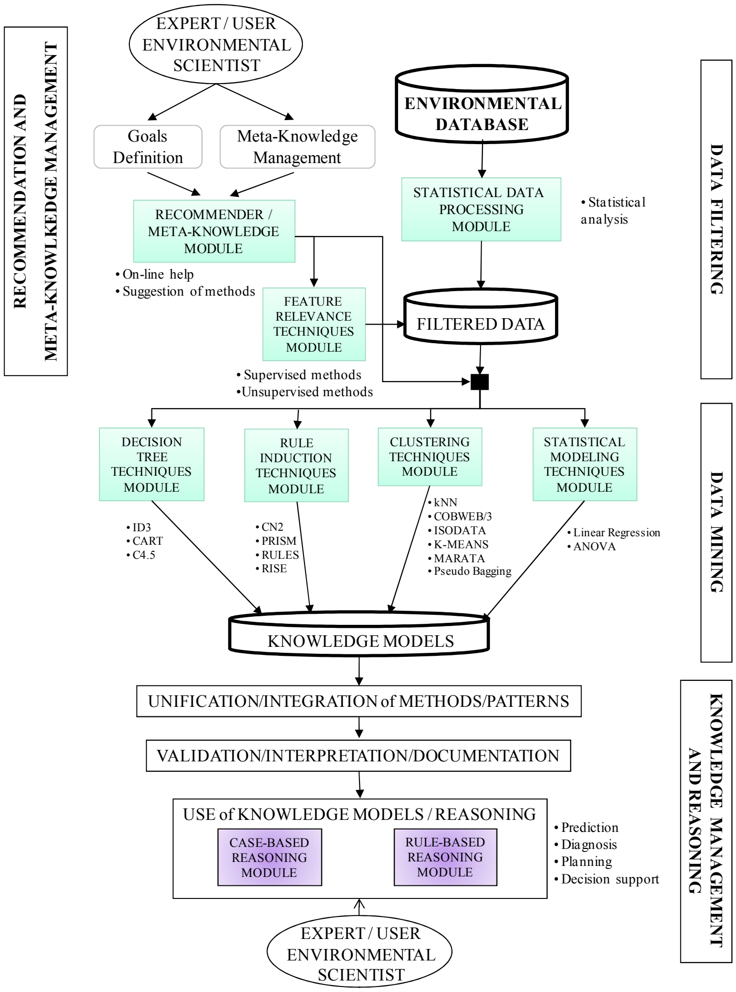
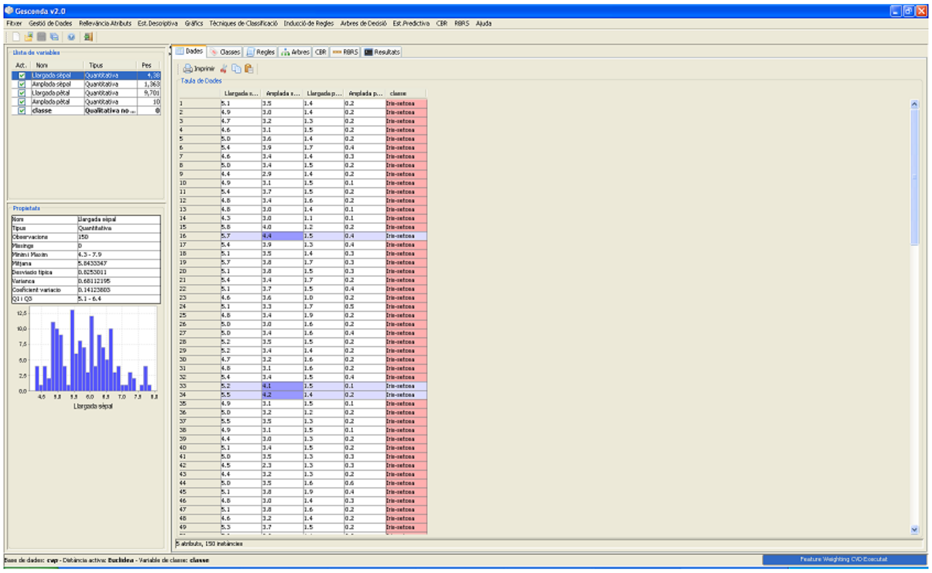

RECOMMENDATION AND META-KNOWLKEDGE MANAGEMENT
EXPERT / USER ENVIRONMENTAL SCIENTIST
Goals Definition
Meta-Knowledge Management
RECOMMENDER / META-KNOWLEDGE MODULE
• On-line help
• Suggestion of methods
ENVIRONMENTAL DATABASE
STATISTICAL DATA PROCESSING MODULE
• Statistical analysis
DATA FILTERING
FEATURE RELEVANCE TECHNIQUES MODULE
FILTERED DATA
• Supervised methods
•Unsupervised methods
DECISION TREE TECHNIQUES MODULE
RULE INDUCTION TECHNIQUES MODULE
CLUSTERING TECHNIQUES MODULE
STATISTICAL MODELING TECHNIQUES MODULE
DATA MINING
• ID3
• CART
• C4.5
• CN2
• PRISM
• RULES
• RISE
• kNN
• COBWEB/3
• ISODATA
• K-MEANS
• MARATA
• Pseudo Bagging
• Linear Regression
• ANOVA
KNOWLEDGE MODELS
UNIFICATION/INTEGRATION of METHODS/PATTERNS
VALIDATION/INTERPRETATION/DOCUMENTATION
USE of KNOWLEDGE MODELS / REASONING
CASE-BASED REASONING MODULE
RULE-BASED REASONING MODULE
• Prediction
• Diagnosis
• Planning
• Decision support
KNOWLEDGE MANAGEMENT AND REASONING
EXPERT / USER ENVIRONMENTAL SCIENTIST

Figure 1. GESCONDA Architecture

  - Statistical one-way analysis;
  - Statistical two-way analysis;
  - Graphical visualization tools;
  - Attribute or Variable transformation
- Recommendation and Meta-Knowledge Management:
  - Problem goal definition;
  - Method suggestion;
  - Parameter setting;
  - Attribute or Variable Meta Knowledge management;
  - Example Meta-knowledge management;
  - Domain theory knowledge elicitation
- Data Mining:
  - Clustering (Machine Learning and Statistical);
  - Decision tree induction;
  - Classification rule induction;
  - Statistical Modelling;
- Knowledge Management and Reasoning:
  - Integration of different knowledge patterns;
  - Validation of the acquired Knowledge pattern;
  - Rule-based reasoning
  - Case-based reasoning

- User interaction.

GESCONDA provides a set of mixed techniques that will be useful to acquire relevant knowledge from environmental systems, through available databases. This knowledge will be used afterwards in the implementation of reliable IEDSS. The portability of the software is provided by a common Java platform. In figure 2 there is a snapshot of the GESCONDA interface.



Act.	Nom	Tipus	Pes
☑	Llargada sèpal	Quantitativa	4,38
☑	Amplada sèpal	Quantitativa	1,363
☑	Llargada pètal	Quantitativa	9,701
☑	Amplada pètal	Quantitativa	10
☑	classe	Qualitativa no ...	0

Nom	Llargada sèpal
Tipus	Quantitativa
Observacions	150
Missings	0
Mínim i Màxim	4.3 - 7.9
Mitjana	5.8433347
Desviació típica	0.8253011
Variança	0.68112195
Coeficient variació	0.14123803
Q1 i Q3	5.1 - 6.4

	Llargada s...	Amplada s...	Llargada p...	Amplada p...	classe
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3.0	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5.0	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5.0	3.4	1.5	0.2	Iris-setosa
9	4.4	2.9	1.4	0.2	Iris-setosa
10	4.9	3.1	1.5	0.1	Iris-setosa
11	5.4	3.7	1.5	0.2	Iris-setosa
12	4.8	3.4	1.6	0.2	Iris-setosa
13	4.8	3.0	1.4	0.1	Iris-setosa
14	4.3	3.0	1.1	0.1	Iris-setosa
15	5.8	4.0	1.2	0.2	Iris-setosa
16	5.7	4.4	1.5	0.4	Iris-setosa
17	5.4	3.9	1.3	0.4	Iris-setosa
18	5.1	3.5	1.4	0.3	Iris-setosa
19	5.7	3.8	1.7	0.3	Iris-setosa
20	5.1	3.8	1.5	0.3	Iris-setosa
21	5.4	3.4	1.7	0.2	Iris-setosa
22	5.1	3.7	1.5	0.4	Iris-setosa
23	4.6	3.6	1.0	0.2	Iris-setosa
24	5.1	3.3	1.7	0.5	Iris-setosa
25	4.8	3.4	1.9	0.2	Iris-setosa
26	5.0	3.0	1.6	0.2	Iris-setosa
27	5.0	3.4	1.6	0.4	Iris-setosa
28	5.2	3.5	1.5	0.2	Iris-setosa
29	5.2	3.4	1.4	0.2	Iris-setosa
30	4.7	3.2	1.6	0.2	Iris-setosa
31	4.8	3.1	1.6	0.2	Iris-setosa
32	5.4	3.4	1.5	0.4	Iris-setosa
33	5.2	4.1	1.5	0.1	Iris-setosa
34	5.5	4.2	1.4	0.2	Iris-setosa
35	4.9	3.1	1.5	0.1	Iris-setosa
36	5.0	3.2	1.2	0.2	Iris-setosa
37	5.5	3.5	1.3	0.2	Iris-setosa
38	4.9	3.1	1.5	0.1	Iris-setosa
39	4.4	3.0	1.3	0.2	Iris-setosa
40	5.1	3.4	1.5	0.2	Iris-setosa
41	5.0	3.5	1.3	0.3	Iris-setosa
42	4.5	2.3	1.3	0.3	Iris-setosa
43	4.4	3.2	1.3	0.2	Iris-setosa
44	5.0	3.5	1.6	0.6	Iris-setosa
45	5.1	3.8	1.9	0.4	Iris-setosa
46	4.8	3.0	1.4	0.3	Iris-setosa
47	5.1	3.8	1.6	0.2	Iris-setosa
48	4.6	3.2	1.4	0.2	Iris-setosa
49	5.3	3.7	1.5	0.2	Iris-setosa

5 atributs, 150 instàncies

Base de dades: **cwp** - Distància activa: **Euclidea** - Variable de classe: **classe**

Feature Weighting CVD Executat

Figure 2. GESCONDA interface

4. RECOMMENDATION AND META-KNOWLEDGE MANAGEMENT LAYER, AND THE DATA FILTERING LAYER

The *recommendation and meta-knowledge layer* includes two modules: the recommender and meta-knowledge module, and the feature relevance module. The former one (see Gibert *et al.* [2010] for further details) let the user to be assisted to select the most suitable methods to be applied, taking into account the goals of the user, and main features of the domain. Also, the meta-information associated with the data can be managed, both from the features or the observations.

The feature relevance module provides GESCONDA with a set of implemented feature weighting algorithms, which determine the weight or relevance of each one of the features describing the data. There are several unsupervised methods and some supervised methods. Among the unsupervised methods implemented there are the Gradient Descent method (GD), the Unsupervised Entropy-Based 1 method (UEB-1) and the Unsupervised Entropy-Based 2 methods (UEB-2). From the supervised methods, in the software there are implemented the Information Gain method (IG), the Projection method (PROJ), the Class Value Distribution method (CVD), and the Entropy-Based Local method (EBL). See Gibert *et al.* [2006] for further details.

The *data filtering layer* wraps the data filtering process, which has to be faced prior to the use of the data mining methods. All this processing is encoded in the statistical data processing module. The statistical data filtering agent is in charge of database management, statistical descriptive analysis, and graphical representations. These tasks provide the whole GESCONDA system with powerful data filtering techniques to prepare the data for later knowledge discovery step.

Database management allows adding a new variable to the database, deleting one variable from the database, and modifying the characteristics of a variable such as its relevance or the range of values. Also, the management of the examples is supported. Some variable transformations such as re-coding and standardization are provided. Different random number distribution generation and different probability distribution generation are supported too.

Descriptive statistical analysis is composed by basic statistical analysis such as computation of mean, standard deviation, median value or correlation coefficient. One-way and two-way analysis of both variables and classes are also provided. Missing value management is also supported. Graphical representations of analysis results are implemented through both one-way plots and two-way plots, as well as histograms or letter plots for class distribution visualization.

5. DATA MINING LAYER

The knowledge discovery layer is the layer joining the modules containing several data mining models, which can be induced from the data. Currently there are four modules developed: the clustering techniques module, the decision tree techniques module, the rule induction module, and the statistical modelling module.

These techniques have been detailed in previous works by Sànchez-Marrè *et al.,*[2004; 2008], and due to lack of space the detailed description here has been skipped.

6. KNOWLEDGE MANAGEMENT AND REASONING LAYER

At this layer is where the new reasoning capabilities provide GESCONDA with analytical, synthetical and predictive skills. Until now GESCONDA was an Intelligent Data Mining and Discovery tool, which was able to induce several knowledge models, but no problem solving skills were given to the user. Now, with the addition of the case-based reasoning module, and the rule-based reasoning module, it has become an actual Intelligent Decision Support Tool, which assists the user to all the steps of the problem solving cycle: diagnosis, planning/solution generation, prediction, and finally decision support. In this section the new two reasoning modules implemented are described.

6.1 Case-Based Reasoning

The Case-Based Reasoning implemented in GESCONDA is conceived to be a Flexible Data-Intensive Case-Based Reasoning. Flexible in terms of extending its functionality and configuration of its functions (see detailed description in Sevilla, [2009]).

This system could be launch interacting with the user at any step (retrieve, reuse, revise and retain) or defining the new case and compute the whole cycle. The main functionalities that are offered are the following:

- Case Structure. Since the data is share it with GESCONDA, it is necessary to define what attributes belongs to the description or solution. I could be introduced by a file or by the interface. Also, there optional attributes for improving the performance:
    - Evaluation. Depending on the type of evaluation a new attribute is created to store the cases that have been used to create the solution of this case.
    - Utility. This attribute pretends to define how useful has been the case in the past. It could be defined as a number, date or frequency, even a combination of those.
- Retrieve. Search for the most similar cases to the new case. Depends on:
    - Indexation of the case base: in future release will be integrated hierarchical structures and self-organizing maps.
    - Threshold: Maximum number of cases to retrieve or/and maximum distance.
    - Distance type and its own parameters (included in GESCONDA)
- Reuse: Adaptation of the retrieved cases solutions to the new solution.
    - Copy: directly transferred.
    - Mean: Mean or mode of the solution.

    - Weighted Mean. The cases are weighted by its distance to the new case or a given attribute (i.e. utility or evaluation attribute)
    - Formula. The user can introduce a formula for the numerical solution attributes
- Revise: Evaluation of the proposed solution
    - By the end-user
    - Approximation of the evaluation assessed by the evaluation of the cases that have been used to create the new one.
    - In testing mode is possible to compare the real solution with the proposed one.
- Retain: Case base learn the proposed case.
    - Defined by the end-user
    - Define a list of conditions that must be satisfied to store the new case in the case base.
        - Each condition is defined in a xml file and has its own java class where it is assessed. The introduction of new conditions is trivial. Already implemented: based on distance, evaluation and a recursive cluster elimination (RCE).
- Execution:
    - Possibility to run more than one case.
        - Individual: one case is solved, from the case base or defined by the user.
        - Battery: a list of cases is solved, from the case base or a file.
    - Mode:
        - Normal
        - Testing: creates a new attributes solution in order to compare with the existing ones and compute the percentage of success or error measures for numerical solution attributes

6.2 Rule-Based Reasoning

The module is focused on a rule inference engine with its main abilities: forward chaining and backward chaining. Some research task must be undertaken to propose the best and more efficient implementation of the Knowledge Bases and the Data Base to get a fast, modular and reliable inference engine. The functionalities can be grouped in:

- Design of the Rule-Based Reasoning System (RBRS)
    - Data Base/Fact Base
    - Knowledge Base
        - Management of Modules
        - Management of Rules
            - Import rules from a file (plain, CLIPS) or from other GESCONDA methods (Classification rules, Decision trees, etc.)
        - Management of Meta-rules
    - Inference Engine
        - Forward chaining: deductive reasoning from data to the goals
        - Backward chaining: validation of a concrete goal/s with the available data
- Execution:
    - Interactive: solves a single problem
    - Batch: solving several problems, simulating a process of the real world.

6.2.1 Inference engine

Forward and backward chaining are the main methods with inference rules and logical implications. Forward chaining implements the logical *modus ponens* inference. Backward chaining finds the most plausible explanations for observed data.

Forward chaining

Forward chaining starts with the available data and uses inference rules to extract more data until a goal is reached. The inference engine can also be executed without a goal in order to get new information from the initial facts base.

An inference engine using forward chaining searches the inference rules until it finds one where the antecedent is known to be true. When found it can conclude, or infer, the consequent, resulting in the addition of new information to its data. Inference engines will iterate through this process until a goal is reached.

Because the data determines which rules are selected and used, this method is called data-driven, in contrast to goal-driven backward chaining inference.

One of the advantages of forward-chaining over backward-chaining is that the reception of new data can trigger new inferences, which makes the engine better suited to dynamic situations in which conditions are likely to change.

In order to do a more optimal and useful implementation, given a state of the facts base and one of the knowledge base, there are several rules to be applied and therefore, a conflict resolution strategy have to be applied. The following strategies are implemented, but could be easily extended:

- Depth-first, where each new rule is added to the begging of the agenda.
- Breadth-first, where each new rule is added to the ending of the agenda.
- Random, where each new rule is added in a random position of the agenda.

In addition there is an option that allows the activation or not of a rule only can be executed once in the inference process.

Backward Chaining

Backward chaining (or backward reasoning) is an inference method used in automated theorem proving, proof assistants and other artificial intelligence applications. Backward chaining is implemented in logic programming by *Selective Linear Definite* clause resolution. This method starts with a list of goals (or a hypothesis) and works backwards from the consequent to the antecedent to see if there is data available that will support any of these consequents. An inference engine using backward chaining would search the inference rules until it finds one which has a consequent that matches a desired goal. If the antecedent of that rule is not known to be true, then it is added to the list of goals. Because the list of goals determines which rules are selected and used, this method is called goal-driven.

7. AN APPLICATION TO SET-UP AN IEDSS FOR A WASTEWATER TREATMENT PLANT

In order to show the GESCONDA added utility of new reasoning capabilities to set-up an IEDSS, the use of rule-based reasoning will be described. Due to lack of space the use of case-based reasoning is detailed in Sevilla and Sànchez-Marrè [2010].

Let us suppose that a wastewater treatment plant manager wants to set-up an intelligent decision support system to help herself/himself to make the appropriate decision at each monitoring time cycle over the system. This kind of system should be able to evaluate the operating state of the plant, to make the most accurate diagnosis of the operating state of the plant, and finally propose several alternatives to the plant manager. The plant manager could then make the most appropriate decision based on the evaluation of the possible alternatives. One possibility to get a knowledge model is to implement it through a knowledge base, and use rule-based reasoning as the main reasoning mechanism to make the diagnosis step, and afterwards, to generate the possible solutions.

The main goal of wastewater treatment plants is to guarantee the outflow water quality referred to certain legal requirements, in order to restore the natural environmental balance, which is disturbed by industry wastes or domestic wastewaters.

The process used to achieve this goal is highly complex; on the one hand, because of the intrinsic features of wastewater treatment processes; on the other hand, because of the bad consequences of an incorrect management of the plant.

One of the most commonly used wastewater treatment is based in the activated sludge technology. A very brief description of the process in this kind of plants is presented: the waste water flows sequentially through three processes which are commonly known as pre-treatment, primary and secondary. Figure 3 depicts its general structure: (i): In the pre-treatment, an initial separation of gross solids, oils and greases from wastewater is performed. (ii) Primary treatment consists of leaving the wastewater in a primary settler for some hours. Suspended solids will deposit down the settler and could be removed from the water. (iii) Secondary treatment occurs inside a biological reactor. A population of microorganisms (biomass) degrades the organic matter solved in the wastewater. A secondary settler is used to separate the treated water from the biomass. The primary and secondary settler outputs (solids and biomass) produce a kind of mud which is the input of another set of processes in the WWTP called sludge line.

Figure 3 shows where all the measures are taken along the plant. This set is heterogeneous and usually there are missing values and the sensors may provide noisy data.

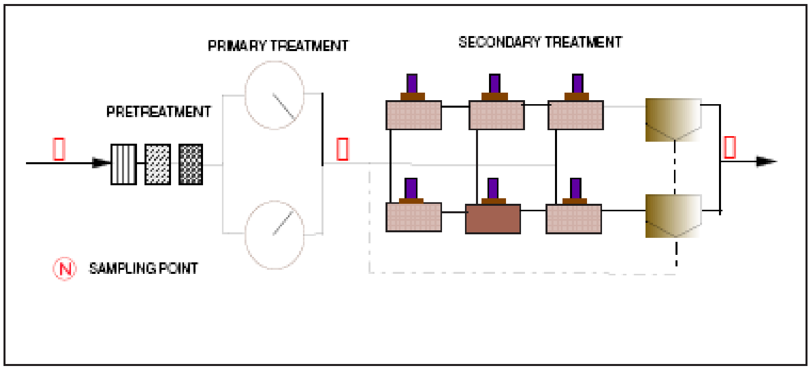


Figure 3. Wastewater treatment plant chart

To set-up this kind of IEDSS let us suppose that there are real data available to be analyzed to discover some knowledge in it, which will be adjusted by the expert knowledge of our plant manager. The experimentation was done with data coming from one wastewater treatment plant in Catalonia. Initially the sample was composed by 303 observations taken from November 1999 to October 2000. Each observation refers to a daily mean, and it is identified by the date itself.

The state of the plant is described through a set of 12 state variables (measured at the entrance, after the primary treatment, within the biological reactor, and at the exit) considered the more relevant upon expert's opinions, plus 3 actuation variables (waste sludge flow, recirculation flow, oxygen inflow). Also, there is one variable expressing a qualitative state of the plant (foaming, rising, bulking, etc.), which was labelled by the experts. There were 19 different labels.

Initially the data were pre-processed using the facilities of GESCONDA software at the data filtering layer (missing values treatment, descriptive statistical analysis of variables, graphical representations, etc.). After the pre-processing step, only 99 observations were kept, because some others had about 60% of missing values. That would be a problem, because for some of the 19 different diagnosis labels, only one or two examples existed, and of course, the generalization power of the rule induction methods will be very poor.

Then the feature relevance module implemented in the recommendation and meta-knowledge layer was used. Several feature weighting methods (IG, CVD, PROJ, etc.) were used to estimate the relevance (weight) of the variables. Finally the CVD method was used.

After that, the data mining step was undertaken using several methods implemented at the data mining layer of GESCONDA. Some classification rule induction methods were used:

PRISM, which obtained 93 very specific classification rules, RISE, which obtained 95 very specific rules and CN2 that only got 3 more general rules. To do that, some discretization methods available at the data filtering layer were used. The rules obtained were very specific because there was a lack of examples of many of the labels.
Also, some decision tree induction methods were used from those available at the data mining layer of GESCONDA. One of the most compact and efficient tree was obtained using C4.5 method. With it 56 rules could be obtained.
In all this data mining process, the plant manager, as a qualified expert was interacting, analysing, and evaluating the results obtained.

Thus, finally a set of inference rules for diagnosing the appropriate label of new operational plant states was obtained. This knowledge model should be the basis of a Knowledge Base. Also, afterwards, some other rules could be discovered to generate the appropriate solutions (setting of the parameter variables) to maintain or restore the normal operating state of the plant.

An example of an induced rule, using PRISM method was:

```
(defrule Rule13
    (?IVF_discret 194,4<IVF<=271,3)
    (?DQO-S_discret 119,3<DQO-S<=177,7)
    (?F:M_discret 0,1<F:M<=0,3)
    (?SS-E_discret 90,0<=SS-E<=236,0)
    (?SS-P_discret 105,3<SS-P<=160,7)
    (?Q-E_discret 16557,3<Q-E<=22014,7)
    (?DQO-E_discret 501,3<DQO-E<=790,7)
    (?DQO-P_discret 559,3<DQO-P<=712,7)
    (?OD_discret 0,8<=OD<=2,7)
    (?SSLM_discret 2470,0<=SSLM<=4314,7)
    (?TRC_discret 2,0<=TRC<=9,6)
    (?SS-S_discret 10,0<=SS-S<=42,3)
    =>
    (assert (Diagnosis Rising)))
```

Most available knowledge discovery tools or Data analysis tools end their functionality here. The added value of the evolution of GESCONDA is that now a rule-based reasoning inference engine is available, and that the rules obtained in the data mining layer can directly be exported to the Knowledge management and reasoning layer. This means that with a very minimal effort, a rule-based inference mechanism can be set-up almost automatically. The knowledge base can be directly transferred. Thus, we have used the 93 rules obtained from the PRISM rule induction method. Moreover, also a case-based reasoning inference engine could have been adopted, and in a near future the combination of both reasoning mechanisms could be possible. This way, the GESCONDA tool covers all cycle of the development of an Intelligent Decision Support System, and in particular for Intelligent Environmental Decision Support Systems, and not only the data mining step, as most of the available software tools.
The rule-based IDSS was tested to diagnose the operational state of other unknown observations, and the accuracy of the diagnosis was around 79%. Afterwards some other control rules were derived to guide the process to safe operational states.

Therefore, the new GESCONDA software with its new reasoning abilities will be a useful shell for developing Intelligent Environmental Decision Support Systems.

8. CONCLUSIONS

The GESCONDA system, which was designed at the beginning, as a Data Mining Tool is becoming an Intelligent Decision Support Tool. Two new reasoning mechanisms like case-based reasoning and rule-based reasoning have been developed and integrated within the tool. These reasoning abilities can provide final users with problem solving, planning, and prediction skills, which make GESCONDA tool a valuable IDSS tool.

In the near future, several new modules will be developed as an artificial neural network module, a support vector machine module, and an evolutionary computation module, and new statistical methods such as PCA, time series, logistic regression, etc., will be integrated too. In addition, the system will be focused, with a high emphasis, on specific 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

The great advantage of GESCONDA in front of other software tools is that in addition to the knowledge discovery process, including data preprocessing and data mining steps, the knowledge management and knowledge model use through reasoning abilities is also possible, converting the software in a valuable tool for the deployment of real IEDSS.

REFERENCES

Gibert, K., M. Sànchez-Marrè, M. and Codina V. Choosing the Right Data Mining Technique: Classification of Methods and Intelligent Recommendation. Procc. of 5th International Environmental Modelling and Software 2010. In press. Ottawa, Canada, July 2010.

Gibert K., Sànchez-Marrè M. and Comas J. The impact of feature weighting in environmental class discovery using GESCONDA. *In 5th ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence (BESAI'2006)*, 5-1:5-10, at ECAI'2006. Riva del Garda, Italy, 2006.

Haagsma, I.G., Johanns, R.D., 1994. Decision support systems: an integrated approach. In: Zannetti, P. (Ed.), Environmental Systems, vol. II., pp. 205–212.

Sànchez-Marrè M. and Gibert K. GESCONDA: from Environmental Data Mining to Environmental Decision Support. 4th Int. Environmental Modelling and Software Society Conference. iEMSs'2008 Procceedings, pp. 1967-1970. Barcelona, 2008.

Sànchez-Marrè M, Gibert K. and Rodríguez-Roda I. (2004). GESCONDA: A tool for Knowledge Discovery and Data Mining in Environmental Databases. In e-Environment: Progress and Challenge. Series on Research on Computing Science 11, 348-364. CIC, IPN, México.

Shi, Z., Y. Huang, H. Qing, X. Lida, S. Liu, L. Qin, Z. Jia, J. Li, H. Huang and L. Zhao. MSMiner--a developing platform for OLAP. *Decision Support Systems* 42, 2016-2028, 2007.

Sevilla, B. A Case-Based Reasoning Shell in a Intelligent Data Analysis System. MSc. Thesis. Universitat Politècnica de Catalunya, 2009.

Sevilla, B. and Sànchez-Marrè, M. Providing Intelligent Decision Support Systems with Flexible Data-Intensive Case-Based Reasoning. Procc. of 5th International Environmental Modelling and Software 2010. In press. Ottawa, Canada, July 2010.