Intelligent Decision Support Systems

(Part XIII - CONCLUSIONS, IDSS TRENDS, ADVANCED TOPICS, OPEN CHALLENGES AND RESOURCES)

Miquel Sànchez-Marrè

Intelligent Data Science and Artificial Intelligence Research Centre (IDEAI-UPC)
Knowledge Engineering and Machine Learning Group (KEMLG-UPC)

Computer Science Dept.
Universitat Politècnica de Catalunya · Barcelona Tech

miquel@cs.upc.edu
http://www.cs.upc.edu/~miquel

Course 2020/2021

https://kemlg.upc.edu
PART 14 – CONCLUSIONS, IDSS TRENDS, ADVANCED TOPICS, OPEN CHALLENGES AND RESOURCES

https://kemlg.upc.edu
CONCLUSIONS
Conclusions

- Reliable IDSS for a real-world complex and ill-structured domains can be constructed
- Multi-knowledge and cooperation of several problem solving techniques (AI, statistical, numerical, control models)
- Data mining is a key step to get inductive models from historical data
- Capturing expert knowledge to get a central corpus (expert-based models)
- Capturing experiential knowledge about either normal situations, prototypical situations or idiosyncratic situations
- Dynamic learning environments
- Portability to similar real-world processes
- Incorporating predictive knowledge to offer prognosis skills
- Effective Decision-Support for real problems can be achieved with IDSS
IDSS TRENDS IN PROCESS CONTROL

https://kemlg.upc.edu
IDSS for Process Control Evolution

- Task
  - Integrated Supervision & Management
  - Single Supervision & Management
  - Intelligent Control Techniques
  - Mathematical Control
  - Diagnosis
  - Design

- Time
  - 70s
  - 80s
  - 90s
  - 2000
  - 2010
  - 2020

- IDSS
  - Interoperable IDSS
  - Multi-Agent IDSS
  - Integrated Intelligent Management Techniques (IDSS)
  - Isolated AI IDSS
  - Isolated Intelligent Control Techniques
  - Intelligent Design techniques
  - Optimal/Predictive Control Techniques
  - Instrumentation & Monitoring Tools

© Miquel Sànchez i Marrè, KEMLG-IDEAI, 2020
ADVANCED TOPICS

https://kemlg.upc.edu
Conceptual Components of an IDSS

- ARTIFICIAL INTELLIGENCE TECHNIQUES
- STATISTICAL / NUMERICAL METHODS
  - UNCERTAINTY MANAGEMENT
  - TEMPORAL REASONING
- INTELLIGENT DECISION SUPPORT SYSTEMS
  - VALIDATION OF IDSS
- GEOGRAPHICAL INFORMATION SYSTEMS / COMPLEMENTARY SYSTEMS
- ENVIRONMENTAL / HEALTH / MANAGEMENT / BUSINESS ONTOLOGIES
- ECONOMICAL COSTS
Advanced Topics

- A common framework/architecture for IDSS development
- Uncertainty Modelling and Management
- Spatial Reasoning
- Temporal Reasoning
A COMMON FRAMEWORK/ARCHITECTURE FOR IDSS DEVELOPMENT

https://kemlg.upc.edu
Dynamic Intelligent Decision Support System (IDSS)

Adapted from [Sànchez-Marrè et al., 2000]
Cognitive-oriented approach for IDSS development (2)
Adapted from [Sànchez-Marrè, 2008]
UNCERTAINTY MODELLING & MANAGEMENT

https://kemlg.upc.edu
Uncertainty Models’ Classification

- **Uncertainty Models**
  - **Probabilistic Models**
    - (Stats) Pure Probabilistic Model
  - **Near-Probabilistic Models**
    - (AI) Certainty Factor Method [MYCIN]
  - **Evidential Model**
    - (AI) Evidence Theory [Dempster-Shafer]
  - **Possibilistic Model**
    - (AI) Possibility Theory Fuzzy Logic [Zadeh]

(AI&Stats) Bayesian Network Model [Pearl]
(AI) Subjective Bayesian Method [PROSPECTOR]
Uncertainty Modelling and Management

- **Model-driven**
  - Expert-based Models
    - Inference rules
      - Fuzzy Logic / Uncertainty factor model

- **Data-driven**
  - Associative Models
    - Bayesian Networks
      - Probabilities
SPATIAL REASONING

https://kemlg.upc.edu
Spatial Reasoning

- Vision: a computational investigation into the human representation and processing of visual information [Marr, 1982]
- Spatial Planning: a configuration space approach [Lozano-Pérez, 1983]
- Representing space for practical reasoning [Fleck, 1987]
- Spatial Reasoning based on Allen’s Temporal Logic [Güsgen, 1989]
- *Qualitative Spatial Reasoning* [Freksa, 1991]
- Cognitive and Linguistics aspects of Geographic Space [Mark & Frank, 1991]
- *Qualitative Spatial Reasoning* about Distances and Directions in geographic Space [Frank, 1992]
- A spatial logic based on regions and connection [Randell *et al.*, 1992]
- A Logic Framework for Reasoning about Space [Lieu, 1993]
- Towards an Integrated Logic of Space, Time and Motion [Galton, 1993]
- *Qualitative representation* of Spatial knowledge [Hernández, 2004]
Prevention of floods

Distance to Radar

Meteorological Radar

Vertical Profile of reflectivity

Different angle elevation

Height

$z_{h1}$

$z_{h2}$

$z_{h3}$

$z_{h4}$

$z_{h5}$

$h_1$

$h_2$

$h_3$

$h_4$

$h_5$
Spatial Neighbourhood Relationship

Reflectivity ($z$): meteorological variable determined by the number and size of the particles presents in a sample volume

As the rainrate increases, the reflectivity increases:

<table>
<thead>
<tr>
<th>Rainrate (in/h)</th>
<th>Reflectivity (dBz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>29.5</td>
</tr>
<tr>
<td>0.25</td>
<td>35.9</td>
</tr>
<tr>
<td>0.5</td>
<td>40.7</td>
</tr>
<tr>
<td>1.25</td>
<td>47.0</td>
</tr>
<tr>
<td>2.5</td>
<td>51.9</td>
</tr>
<tr>
<td>4.0</td>
<td>55.1</td>
</tr>
</tbody>
</table>

© Miquel Sànchez i Marrè, KEMLG-IDEAI, 2020
TEMPORAL REASONING

https://kemlg.upc.edu
The problem

- **Continuous or dynamic or temporal domains**
  - A large amount of new valuable experiences are continuously generated
  - The current state or situation of the domain *depends on previous temporal states or situations* of the domain
  - States have *multiple diagnoses*

- **Examples**
  - Monitoring and on-line control of power stations
  - Wastewater treatment plants control
  - Jet plane control
  - Monitoring of patients in an intensive care unit
  - Diagnosis and/or prognosis and cure of some medical diseases
  - Forecasting of some meteorological or seismic phenomena
  - Autonomous robot navigation control
Temporal Reasoning

- Temporal Interval Logic model [Allen, 1983]
- Dynamic Bayesian Networks (DBN) [Dean & Kanazawa, 1989; Gahum et al., 1991; Nicholson & Brady, 1992; Kjaerulff, 1992]
- Dynamic ANNs
  - Long Short-Term Memory (LSTMs), a kind of Artificial RNN [Hochreiter & Schmidhuber, 1997]
- Episode-Based Reasoning (EBR) [Sànchez-Marrè et al., 2005]
- Incremental Machine Learning Techniques
  - Data Stream Mining [Agarwal, 2007; Gama, 2010]
EBR Approach

- *Episode-Based Reasoning* is based on the abstraction of temporal sequences of cases, which are named as *episodes*.

- Open questions:
  - How to determine the length of an episode?
  - How to represent the episodes, taking into account that they could be overlapping?
  - How to represent the isolated cases?
  - How to relate them to form episodes?
  - How to undertake the episode retrieval?
  - How to evaluate the similarity between temporal episodes of cases?
  - How to continually learn and solve new episodes?
OPEN CHALLENGES
Open Challenges

- IDSS Validation
- Integration of Models: *Interoperation* of Artificial Intelligence/statistic/mathematical models
VALIDATION OF AN IDSS

https://kemlg.upc.edu
Validation of IDSS

- **Validation** of an IDSS: the design of a *set of test cases* or *scenarios* which ensure good performance of the system.
- Good performance can be identified with the capacity of the system to provide the *right recommendation* in front of a certain scenario.
- Evaluation has to be done for a rather specific application domain.
Design of a Standardized Validation protocol

- **Identify the components of the IDSS** as well as their characteristics:
  - Models available, data sources – sensors, laboratories, observations, opinions, etc.- and data quality, knowledge based or soft computing reasoning, learning capability, user profile, system autonomy, open/limited situations faced, etc.

- **Identify the tasks performed by the IDSS**. There are two main kinds of IDSS that should be distinguished regarding its functionality:
  - Those which are controlling / supervising a process in real-time (or almost real-time)
  - Those that give punctual off-line support to decision-making.
Kind of IDSS Tasks

- Three main processes or tasks can be identified:
  - **Diagnosis/Situation Assessment**: based on observation, and oriented to determine “what is going on?”
  - **Synthesis/Generation of alternatives**: based on a specification of goals, determine “what can be done to achieve the goals given a certain diagnoses?”
  - **Decision Support/Recommendation/Therapy proposal**: Prognosis for each possible alternative according to some forecasting model (“what if scenario analysis”) and recommendation of best alternatives.
Two-level Validation process of an IDSS

- So, a general framework could be establishing different evaluation steps that should be fulfilled based on each IDSS particularities:
  - **Structural evaluation**: related to the components of the system and their interaction
  - **Functional evaluation**: Evaluate the good performance of every task involved in the IDSS.
IDSS INTEROPERATION

https://kemlg.upc.edu
IDSS Modelling

- **IDSSs use models of the real world** to try to get an insight of the **behaviour and evolution of the real system**.

- A **model** is a description of a system, usually a simplified description (abstraction) less complex than the actual system.

- **Kind of models:**
  - **Mechanistic models:** based on an understanding of the behaviour of a system's components, analysing the system from its *first-principles*
    - Set of mathematical formulas and equations (differential equations, etc.)
  - **Empirical models:** based on direct observation (expertise and experience), measurement and extensive data records.
    - Data-driven models (data mining):
      - Mathematical and statistical methods (MLR, PCA, DA, LR, etc.)
      - Intelligent data analysis models (Ass. Rules, Classif. Rules, DT, ANN, CBR, Fuzzy Logic, EvoComp, BayNets, etc.)
    - Model-driven techniques:
      - Expert-based models
      - Model-based reasoning techniques
      - Qualitative reasoning models
      - Multi-Agent Simulation models
IDSS Empirical Models & Interoperability

- Three kind of processes regarding models:

  - Model Production task
    - *Model Producers*: Data mining algorithms

  - Model Use Tasks
    - *Model Visualizers*: Visualizing algorithms
    - *Model Executors*: Interpretation algorithms, which are able to use the model for predictive/discriminatory purposes. Also, they can generate new models, adopting the role of model producers.
Interoperability

- *Interoperability*: “The ability of two or more systems or components to exchange information and to use the information that has been exchanged” [IEEE, 1990]

- Additionally, when the components share a common understanding of the information model behind the data being interchanged, *semantic interoperability* is achieved [Manguinhas, 2010]
IDSS Interoperability
Tools/Techniques for Achieving Interoperability (1)

- Use of a **model interchange format** based on the *Predictive Modelling Markup Language (PMML)*, developed by the Data Mining Group (DMG, http://www.dmg.org).

- PMML uses XML to represent mining models.

- The structure of the models is described by an XML Schema.

- A PMML document is an XML document with a root element of type PMML.

- A PMML document can contain more than one model, and most common data mining models are supported (AssociationModel, RegressionModel, TreeModel, RuleSetModel, NeuralNetwork, ClusteringModel, etc.).
<?xml version="1.0"?>
<PMML version="4.2"
xmlns="http://www.dmg.org/PMML-4_2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

<Header copyright="Example.com"/>
<DataDictionary> ... </DataDictionary>
<RuleSet model ...>
 ... a model ...
</RuleSetModel>
...
</PMML>
Tools/Techniques for Achieving Interoperability (3)

- **Visual workflow** as graphical notations to describe processes. Workflow will include:
  - Data
  - Interchange PMML models
  - Model producers
  - Model executors
  - Solution combiners
  - Current problem data specifications

- jBPM is proposed
  - Open source Business process engine supporting the Business Process Model and Notation (BPMN 2.0) standard
  - Providing a graphical workflow editor
Tools/Techniques for Achieving Interoperability (4)

- IDSS solving process (model production, diagnosis tasks, solution-generation tasks, prognosis tasks) will be described through the workflow

- Direct execution of the workflow or generation of the corresponding software code, which will be minimally modified
Tools/Techniques for Achieving Interoperability (5)

- **Ontologies**
  - An ontology for describing
    - AI models
    - Mathematical/Statistical models
  - To assist the automatic execution of the workflows

- **Multiagent approach** in the prognosis layer
  - Agent-based simulation tool
  - Modelling and evaluation of complex scenarios
  - Use of JADEX as a multi-agent platform
    - JADEX is an add-on of JADE agent platform, following the BDI architecture
The Interoperable IDSS Framework Proposal
Using the Framework in a case Study (1)

- Supervision and management of a Wastewater Treatment Plant (WWTP) based on Activated Sludge Process
- The quality of the effluent water must be always maintained in a good condition to minimise any environmental impact.
- In WWTP operation, several problems frequently appear:
  - Solid separation problems
  - Biological foam
  - Bulking episodes in the bioreactors
  - Overloading derived from storms and heavy rains.
- In WWTP operation usually there are three control variables:
  - Dissolved Oxygen (DO in mg/l) in the bioreactor
  - Recirculation Flow (RF in m³/day)
  - Waste Flow (WF in m³/day)
Using the Framework in a case Study (2)
ADVANCED APPLICATIONS

https://kemlg.upc.edu
Advanced Applications

- Recommender Systems, a special case of IDSS
Recommender Systems

https://kemlg.upc.edu
General Architecture
Main Entities

- **Item**
  - the entity being recommended (products, songs, web pages, services, etc.)
  - let $I = \{i_1, i_2, ..., i_n\}$ be the set of available items

- **User**
  - the entity receiving the recommendations
  - Let $U = \{u_1, u_2, ..., u_k\}$ be the set of users identified

- **Transaction**
  - a recorded interaction between a user and a system
  - Let $T = \{t_1, t_2, ..., t_m\}$ be the set of recorded transactions
User–Item Matrix

- **Utility function**, \( v: U \times I \rightarrow R \), the utility that a user from \( U \) gets by using an item (or performing an action) from \( I \).
- The **utility of an item** can be represented by the *rating* a user assigns to an item, showing the level of the item likeliness for a user.
- These **ratings** can be expressed in the **user-item matrix**:

<table>
<thead>
<tr>
<th>User \ Item</th>
<th>( i_1 )</th>
<th>( i_2 )</th>
<th>\ldots</th>
<th>( i_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 )</td>
<td>( r_{11} )</td>
<td>( r_{12} )</td>
<td>\ldots</td>
<td>( r_{1n} )</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>( r_{21} )</td>
<td>( r_{22} )</td>
<td>\ldots</td>
<td>( r_{2n} )</td>
</tr>
<tr>
<td>( u_k )</td>
<td>( r_{1k} )</td>
<td>( r_{2k} )</td>
<td>\ldots</td>
<td>( r_{nk} )</td>
</tr>
</tbody>
</table>

- **RSs**: recommend the item (action) \( i' \in I \) that maximizes user’s (\( u \)) (predicted) utility under certain circumstances, thus:

\[
\forall i \in I, \quad i' = \text{argmax} \ v(u, i)
\]
Examine **user-item** matrix

<table>
<thead>
<tr>
<th>User \ Item</th>
<th>( i_1 ) ([a_{11}, \ldots, a_{1t}])</th>
<th>( i_2 ) ([a_{21}, \ldots, a_{2t}])</th>
<th>( i_n ) ([a_{n1}, \ldots, a_{nt}])</th>
</tr>
</thead>
<tbody>
<tr>
<td>( u_1 )</td>
<td>( r_{11} )</td>
<td>( r_{21} )</td>
<td>( r_{n1} )</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>( r_{12} )</td>
<td>( r_{22} )</td>
<td>( r_{n2} )</td>
</tr>
<tr>
<td>( u_k )</td>
<td>( r_{1k} )</td>
<td>( r_{2k} )</td>
<td>( r_{nk} )</td>
</tr>
</tbody>
</table>

**Collaborative Filtering**
- Users’ similarity through ratings
- Items’ similarity through attributes

**Content-Based**
- Recommend Items
- Recommend Similar items

Recommend Items **of similar users**
Recommender Systems: an example from Amazon

https://kemlg.upc.edu
Amazon Recommendations (1)

- Recommending new books for a registered user
- Procedure:
  - Store the necessary data about the users of the system and their purchases
    - Users
    - Books
    - CDs
    - Other products
  - Focus especially on *what books (items) they buy*
  - Use the data and information of all registered users through several intelligent techniques to make appropriate recommendations for new books (items)
Amazon Recommendations (2)

- Content-Based Recommendation
Amazon Recommendations (3)

- Association-Based Recommendation (who? Has what?)
- Collaborative-Filtering Recommendation
Amazon Recommendations (4)

- Amazon algorithm: *item-to-item collaborative filtering (CB)*

- Algorithm *item-to-item collaborative filtering (CB):*
  1. Off-line calculation of the similarity between a single product and all related products:
     
     For each item in product catalog, \( I_1 \)
     
     For each customer \( C \) who purchased \( I_1 \)
     
     For each item \( I_2 \) purchased by customer \( C \)
     
     Record that a customer purchased \( I_1 \) and \( I_2 \)
     
     For each item \( I_2 \)
     
     Compute the similarity between \( I_1 \) and \( I_2 \)
  
  2. Match each of the user’s purchased and rated items to similar items (CB)
  
  3. Combine those similar items into a recommendation list

© Miquel Sànchez i Marrè, KEMLG-IDEAI & Amazon, 2020
RESOURCES

https://kemlg.upc.edu
Resources

- Tools
- References
SOFTWARE TOOLS FOR IDSS DEVELOPMENT

https://kemlg.upc.edu
Software Tools for an IDSS Development

- General Tools
- Tools for Data-driven Methods
- Tools for Model-driven Techniques
- General Development Environments
- GESCONDA project
General Tools

- Case-Based Reasoning
  - JColibri framework (https://gaia.fdi.ucm.es/research/colibri/jcolibri)
  - IUCBRF (https://homes.luddy.indiana.edu/leake/iucbrf/index.html)
  - myCBR (http://mycbrr-project.org/)

- Ontologies & Ontological Reasoning
  - Protegé (http://protege.stanford.edu/)
  - Hozo (http://www.hozo.jp/)
  - DogmaModeler (http://www.jarrar.info/Dogmamodeler/index.htm)

- Numerical Computation & Simulation
  - Matlab & Simulink (https://www.mathworks.com)
  - GNU Octave (https://www.gnu.org/software/octave/index)
  - Scilab (https://www.scilab.org/)
Tools for Data-driven Methods

- WEKA
- Rapid Miner Studio
- KNIME Analytics Platform
- KEEL
- Orange
- IBM SPSS Modeler
- SAS Enterprise Miner
Tools for Model-driven Techniques (1)

- Agent-based Simulation Tools
  - Prometheus
  - Jadex
  - Ascape
  - FLAME
  - Janus
  - Netlogo
  - Anylogic

- Expert-based Model Tools
  - CLIPS (C-Language Integrated Production System)
  - Drools (Rule-Based Engine)
  - JESS (Java Expert System Shell)
Tools for Model-driven Techniques (2)

- Model-based Reasoning Tools
  - MINION
  - Gecode
  - Choco solver

- Qualitative Reasoning Tools
  - Garp3
  - GQR (Generic Qualitative Reasoner)
  - Simantics System Dynamics
General Development Environments

- **R System**
  - RStudio desktop IDE

- **Python**
  - PyDev (Python IDE for Eclipse)
  - PyCharm (community edition) IDE by © JetBrains
  - Spyder (Scientific PYthon Development EnviRonment)
R SYSTEM

https://kemlg.upc.edu
R System

- R is a language and environment for statistical computing and graphics.
- It is a GNU project which is similar to the S language and environment which was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues.
- R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible.
- One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. Great care has been taken over the defaults for the minor design choices in graphics, but the user retains full control.
- R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form.
- It compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and MacOS.
The R Environment

- R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes:
  - an effective data handling and storage facility,
  - a suite of operators for calculations on arrays, in particular matrices,
  - a large, coherent, integrated collection of intermediate tools for data analysis,
  - graphical facilities for data analysis and display either on-screen or on hardcopy, and
  - a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities.
RStudio

```r
R version 3.1.1 (2014-07-10) -- "Sock it to Me"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information
and 'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

© Miquel Sànchez i Marrè, KEMLG-IDEAI, 2020
Python

https://kemlg.upc.edu
Python

- Python is a general purpose, multi-paradigm programming language, supporting object-oriented, functional and procedural coding structures.
- Python is an interpreted, and interactive programming language. It can be used both for standalone applications and for scripting applications in many domains.
- Main features:
  - Combining remarkable power with easiness to use
  - Simple and consistent programming language syntax and semantics
  - Cross-platform availability
  - Highly modular: code can be grouped into modules and packages
  - Suited both for rapid prototyping and large-scale programming
  - Supporting object-oriented programming with multiple inheritance
  - Powerful large standard library is included for numeric processing, image manipulation, user interfaces, web scripting, etc.
  - There are interfaces to many system calls and libraries, as well as to various windowing systems.
  - New built-in modules are easily written and compiled in C or C++ or other languages for higher speed.
  - Usable as an extension language for applications written in other languages that need easy-to-use scripting or automation interfaces.
PyCharm
Spyder
GESCONDA: an Intelligent Decision Support Tool

https://kemlg.upc.edu
GESCONDA

- GESCONDA (Sànchez-Marrè et al., 2010)
  [https://kemlg.upc.edu/en/projects/gesconda-1/gesconda]
  - It was conceived as a system for Knowledge Discovery/Data Mining and knowledge management from databases.
  - Capabilities:
    - Data-driven models can be mined from data.
    - Currently supports additional functionalities case-based reasoning, rule-based reasoning and genetic algorithms are provided.
    - It is a 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.
  - Developers: Miquel Sànchez-Marrè, Karina Gibert & other members of KEMLG (IDEAI-UPC)
GESCONDA tool
Data Base Management & Statistical Data Processing

- Database Management
  - Adding variables
  - Deleting variables
  - Recoding variables
  - Standardization of variables
  - Random variable generation
  - Probability values generation

- Statistical Data Filtering
  - Descriptive statistical analysis:
    - one-way
    - two-way
  - Graphical representation
    - TS-Plots
    - Histograms
    - Letter plots
    - Two-way charts
Statistical Analysis
Plots
Feature Relevance module

- Supervised Feature Weighting methods
  - IG
  - Proj
  - CVD
  - EBL

- Unsupervised Feature weighting methods
  - Gradient method
  - UEB-1
  - UEB-2

- Functionalities
  - Weight assignment
  - Weights exportation
Feature Weighting

Determinació automàtica de pesos

Tria l'algoritme i fixa els paràmetres corresponents:

- Algorismes
  - GD (Gradient)
  - UEB-1 (Unsupervised Entropy Based 1)
  - UEB-2 (Unsupervised Entropy Based 2)

- Alfa: 0
- Mu: 0
- Threshold: 1
- Exponent: -5
- Iteracions: 5

Pesos assignats

<table>
<thead>
<tr>
<th>Atributs i pesos</th>
<th>Valors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLV55-B</td>
<td>6.3243303</td>
</tr>
<tr>
<td>DQO-E</td>
<td>6.7626553</td>
</tr>
<tr>
<td>DQO-D</td>
<td>5.8280478</td>
</tr>
<tr>
<td>DBO-E</td>
<td>8.54659</td>
</tr>
<tr>
<td>DBO-D</td>
<td>5.566908</td>
</tr>
<tr>
<td>Q-E</td>
<td>5.087847</td>
</tr>
<tr>
<td>V30-B</td>
<td>3.6127439</td>
</tr>
<tr>
<td>QR-G</td>
<td>9.481031</td>
</tr>
<tr>
<td>PH-E</td>
<td>4.863965</td>
</tr>
<tr>
<td>DQO-S</td>
<td>5.803375</td>
</tr>
<tr>
<td>PH-D</td>
<td>5.577009</td>
</tr>
<tr>
<td>MCRT-B</td>
<td>10.0</td>
</tr>
<tr>
<td>DBO-S</td>
<td>8.025634</td>
</tr>
<tr>
<td>QB-B</td>
<td>7.019821</td>
</tr>
<tr>
<td>SS-E</td>
<td>7.8666644</td>
</tr>
<tr>
<td>SS-D</td>
<td>6.9835567</td>
</tr>
<tr>
<td>SSV-E</td>
<td>8.026958</td>
</tr>
<tr>
<td>SSV-D</td>
<td>7.5566745</td>
</tr>
<tr>
<td>QP-G</td>
<td>5.0625114</td>
</tr>
<tr>
<td>PH-S</td>
<td>3.7629368</td>
</tr>
<tr>
<td>FE-E</td>
<td>6.8153143</td>
</tr>
<tr>
<td>SS-S</td>
<td>9.310967</td>
</tr>
<tr>
<td>QA-G</td>
<td>0.0</td>
</tr>
<tr>
<td>SSV-S</td>
<td>9.303681</td>
</tr>
<tr>
<td>MLSS-B</td>
<td>6.4036493</td>
</tr>
</tbody>
</table>

Acceptar | Exportar a fitxar | Rebutjar
Clustering module

- Clustering methods
  - Nearest Neighbour
  - COBWEB/3
  - Isodata
  - K-means
  - Marata
  - Hybrid clustering based on rules

- Randomised Pseudo-bagging techniques
  - K-means
  - Nearest Neighbour
Clustering plot
Decision Tree Induction module

- Decision tree classifiers
  - ID3
  - C4.5
  - CART

- Functionalities
  - Pruning
  - Validation
C4.5 induced tree
Statistical Modelling module

- Multiple Linear Regression
  - Parameter estimation
  - Evaluation of fitting coefficients
  - Providing tools for graphical residuals analysis
- ANOVA
  - Parameter estimation
  - Evaluation of fitting coefficients
  - Providing tools for graphical residuals analysis
- Logistic Regression (coming soon ...)
- Multivariate Data Analysis (coming soon ...)
Classification Rule Induction module

- **Methods**
  - CN2 method
  - PRISM method
  - RISE method
  - Rules method

- **Functionalities**
  - Rule pruning
  - Validation (simple validation, cross validation)
  - Prediction
  - Rule exportation (CLIPS format, TXT format)
Rule set exported in CLIPS format

(defrule Regla1
  (>= ?Llargada_pètal 1.0)
  (< ?Llargada_pètal 2.18)
  =>
  (assert (classe Iris-setosa)))

(defrule Regla2
  (>= ?Amplada_pètal 0.1)
  (< ?Amplada_pètal 0.58)
  =>
  (assert (classe Iris-setosa)))

(defrule Regla3
  (>= ?Amplada_pètal 1.06)
  (< ?Amplada_pètal 1.54)
  (>= ?Llargada_pètal 3.36)
  (< ?Llargada_pètal 4.54)
  =>
  (assert (classe Iris-versicolor)))

(defrule Regla4
  (>= ?Amplada_pètal 2.02)
  (< ?Amplada_pètal 2.5)
  =>
  (assert (classe Iris-virginica)))

...
Association Rules

- A priori algorithm
- A priori TID algorithm
- A priori Hybrid algorithm
Reasoning

- Case-Based Reasoning
- Rule-Based Reasoning
- Genetic Algorithms
Genetic Algorithm Interface

- **General**
  - No. of Generations: 50
  - Population size: Number of Retrieved Cases
  - Initial Population type: Random
  - Encoding scheme: Real
  - Termination criteria: Run for specified num. of generations

- **Fitness function**
  - min[distance(NewCase, retrievedCase_i)]

- **Selection**
  - Selection criteria: Tournament
  - Tournament size: 5
  - Selection type: Elist
  - No. of elite children: 2

- **Crossover**
  - Crossover type: Single point
  - Crossover rate: 0.6

- **Mutation**
  - Mutation: Real mutator
  - Mutation rate: 0.1
Conclusions

- Capturing knowledge implicit in databases to get a central corpus for IDSS/IS is possible
- Reliable IDSS/IS for a real-world complex and ill-structured domains can be constructed
- GESCONDA is a promising Intelligent Data Analysis System for building real Intelligent Decision-Support Systems
- GESCONDA integrates both statistical and artificial intelligence data mining techniques
- Future recommendation and meta-knowledge tools will improve its usability
References (1)

General References

References (2)

References (3)

References (4)

References (5)

References (6)

- IDSS Applications
  
  
  
  
References (7)


References (8)


References (9)


References (10)


References (11)


References (12)


References (13)

References (14)


References (15)


References (16)


References (17)


References (18)


Internet Resources

https://kemlg.upc.edu
Internet Resources

- Catalan Statistics Institute (Institut d'Estadística de Catalunya)
  http://www.idescat.net
- European Statistics Office
  http://www.europa.eu.int/comm/eurostat
- UCI Machine Learning Repository
  http://archive.ics.uci.edu/ml/
- Knowledge Discovery Network of Excellence
  http://www.kdnet.org/
- Data Mining Resources
  http://www.scd.ucar.edu/hps/GROUPS/dm/dm.html
- Data Mining Sites
  http://www.cacs.louisiana.edu/~arlab/repos/sortedSites.html
- GESCONDA software
  http://wwwlsi.upc.edu/~webia/KEMLG/projects/gesconda.html
- WEKA Software
  http://www.cs.waikato.ac.nz/~ml/
- R Software
  http://www.r-project.org/
- DAVIS Software
  http://stat.skku.ac.kr/myuh/DAVIS.html
Intelligent Data Science and Artificial Intelligence (IDEAI-UPC)

Miquel Sànchez-Marrè
miquel@cs.upc.edu

Knowledge Engineering and Machine Learning Group
UNIVERSITAT POLITÈCNICA DE CATALUNYA

https://kemlg.upc.edu