Intelligent System Project

(Part IV - DEVELOPMENT OF AN ISP)

Miquel Sànchez i Marrè
miquel@cs.upc.edu

http://kemlg.upc.edu/menu1/miquel-sanchez-i-marre

Course 2015/2016

https://kemlg.upc.edu

Knowledge Engineering and Machine Learning Group
UNIVERSITAT POLITÈCNICA DE CATALUNYA
PART 4 – DEVELOPMENT OF AN INTELLIGENT SYSTEM PROJECT

https://kemlg.upc.edu
Development of an ISP

- Data/Information Extraction
- Knowledge/Ontological Analysis
- Data Mining & Knowledge Acquisition Process
- Planning and selection of Intelligent/Statistical/Mathematical Methods/Techniques
- Construction of Models and implementation of Techniques
- Module Integration
- Validation of Models/Techniques. Comparison of Techniques
- Proposed Solution
Data/Information Extraction

- Planning for data collection
  - Sampling
- Data collection
- Purification and filtering data
  - Error Detection and Repairing
  - Outlier Management
  - Missing Values
- Meta-Data gathering
- Data Relevance
Knowledge/Ontological Analysis

- Knowledge acquisition is hard
- AI/Computer scientists are not especially prepared
- Ontology formation means to study and learn a new domain and new tasks

Methods:
- Literature revision
- Expert interviews:
  - Questions and answers for basic principles
  - Explicit scenarios or use cases
- Automatic Knowledge acquisition from data
  - Use of Data Mining/Machine Learning methods
Data Mining & Knowledge Acquisition Process (1)

- Several descriptive analyses
  - Visualization
  - Histogram / Bar chart
  - TS plots
  - Bivariate diagrams

- Obtaining new variables
  - Recoding
  - Formation of response/target variable/s
    - Y=f(…)

- Distribution of response/target variable/s
Data Mining & Knowledge Acquisition Process (2)

- Variable/attribute relevance
  - Bivariate diagrams
  - Contingency tables
  - Boxplots
  - Techniques of *feature selection* and *feature weighting*

- Variable selection: determination of the explanatory variables
  - \( Y = f(X_1, X_2, \ldots, X_n) \)
Planning and selection of Intelligent / Statistical / Mathematical Methods (1)

- Depending on the *objective* and *sub-objectives* of the problem and the data, the methods are chosen.
- Some methods may involve:
  - Additional data transformations
  - Assumptions / hypotheses
- Different types of models
  - Statistical models
  - Artificial Intelligence Models
  - Hybrid Statistical / AI Models
  - Optimization and Operations Research Models
Planning and selection of Intelligent / Statistical / Mathematical Methods (2)

- Statistical Techniques
  - Linear Models: simple regression, multiple regression
  - Time Series Models (AR, MA, ARIMA)
  - Component Principal Analysis (CPA) / Discriminant Analysis (DA)

- Artificial Intelligence Techniques
  - Decision Trees
  - Classification Rules
  - Association Rules
  - Clustering
  - Instance-Based Learning (IBL, CBR)
  - Connectionist Approach (Artificial Neural Networks)
  - Evolutionary Computation (Genetic Algorithms, Genetic Programming)
Planning and selection of Intelligent / Statistical / Mathematical Methods (3)

- AI & Stats Techniques
  - Regression Trees
  - Model Trees
  - Probabilistic/Belief/Bayesian Networks
  - Support Vector Machines (SVMs)
Model Classification (1)

Data Mining Models

Models with No response variable / Unsupervised models
- Descriptive Models
  - Conceptual Clustering Self Organising Maps (SOMs)
  - Statistical clustering
  - Clustering based on rules (CIBR)

Models with response variable / Supervised models
- Associative Models
  - Association Rules Model-based Reasoning Qualitative Reasoning
  - Principal Component Analysis (PCA)
  - Simple Correspondence Analysis (SCA)
  - Multiple Correspondence Analysis (MCA)
  - Bayesian networks (BayNet)

- Discriminant Models
  - Case-based models
    - Case-based Classifier (CBRClas)
  - Rule-based models
    - Rule-based Classifiers Decision Trees
  - Bayesian models
    - Naïve Bayes Classifier

- Predictive Models
  - Connexionist models (NeuralNet)
  - Case-based Predictor (CBRPred)
  - Evolutionary Computing (GAs)
  - Swarm Intelligence
  - Linear Regression (LR)
  - Multiple Linear Regression (MLR)
  - Analysis of Variance (ANOVA)
  - Generalized Linear Models (GLM)
  - Time Series (TS)

Models with response variable / Supervised models
- Rule-based Classifiers
  - Decision Trees

Models with no response variable / Unsupervised models
- Case-based Classifiers
  - Regression Trees
  - Model Trees
  - Support Vector Machines (SVM)
Model Classification (2)

UNCERTAINTY MODELS

PROBABILISTIC MODELS
  (Stats) Pure Probabilistic Model
  (AI &Stats) Bayesian Network Model [Pearl]

NEAR-PROBABILISTIC MODELS
  (AI) Certainty Factor Method [MYCIN]
  (AI) Subjective Bayesian Method [PROSPECTOR]

EVIDENTIAL MODEL
  (AI) Evidence Theory [Dempster-Shafer]

POSSIBILISTIC MODEL
  (AI) Possibility Theory Fuzzy Logic [Zadeh]
Construction of Models and Implementation of Techniques

● Using different software tools
  ■ Statistics tools
    ◆ MINITAB
    ◆ SPSS
    ◆ SAS
    ◆ SPAD
    ◆ SYSTEM-R
  ■ Optimization and Operation Research tools
  ■ Artificial Intelligence tools
    ◆ WEKA
    ◆ Matlab
    ◆ GESCONDA

● Using Programming languages
  ■ Java
  ■ C++
  ■ Python
  ■ CommonLisp
  ■ ...
Module Integration

● Manual Integration of Models

● Model Interoperability
  ■ Interoperability is the ability of two or more systems or components *to exchange information* and *to use the information* that has been exchanged.

● Semantic Interoperability of Models
  ■ Additionally, when the components share a *common understanding* of the information model behind the data being interchanged, semantic interoperability is achieved.
Validation of Models/Techniques. Comparison of Techniques

OBJECTIVES

- Accuracy and reliability of the models obtained
- Scalability / generalization of the models
- Interpretability of models
- Flexibility and ease of use of the models

METHODS

- Validation
  - Simple / Cross-Validation
  - Random / Stratified
- Confusion matrices
- Table of classification errors
- ROC Curves
  - Gini index
Proposed Solution

- Method/s and model/s chosen
- Justification of the choice
- Estimation accuracy/goodness of the method(s)/model(s) chosen
Examples of Intelligent System Projects

- Analysis of the internal behaviour of Internet web pages regarding to web links design and web surfing efficiency to improve the quality of WWW
- Analysis and prediction of a company customers’ loyalty
- An intelligent urban planning transport system
- An intelligent system to propose the best wastewater treatment technology for a given town
- An intelligent recommender system for market basket analysis
Miquel Sànchez i Marrè
(miquel@cs.upc.edu)

http://kemlg.upc.edu/