Machine Learning Group

Department of Computer Science
Université Libre de Bruxelles

Member of the

- *CIL - Computational and Intelligence Learning Doctoral School*
- *CINBIOS - Bioinformatics and Biomodelling Center.*

WEB Site: http://www.ulb.ac.be/di/mlg

Coordinator: Gianluca BONTEMPI

gbonte@ulb.ac.be

ULB Machine learning group

- 1 faculty, 7 PhD candidates, 3 postdocs, 1 engineer
- Research topics:
  - **Geographical data mining** (Partners: ISYS at UCL, Louvain-La-Neuve and HILS at KUL, Leuven, KDD Lab, ISTI CNR, Italy),
  - **Fraud detection** (Partners: IGEAT and SLN at ULB, SIC at Royal Military Academy),
  - **Data processing in wireless sensor networks** (Partner: IRIDIA, Voice Insight),
  - **Bioinformatics – Gene expression and cancer detection** (Partners: Jules Bordet Institute, Molecular Biology and Chemistry departments at ULB),
  - **Computer-aided medicine** (Partners: University Hospitals of Wallon Region, Mexys SA, Arificial intelligence laboratory in Milano),
  - **Theoretical research**: Local learning, feature selection, model selection.
- Facilities:
  - Cluster of 16 computers,
  - 75 ultra low power wireless sensors,
  - Lego robotics lab (10 Mindstorms and 2 NXT kits, One Surveyor mobile robot).
- Website: www.ulb.ac.be/di/mlg

Projet ARMURS

Automatic Recognition for Map Update by Remote Sensing

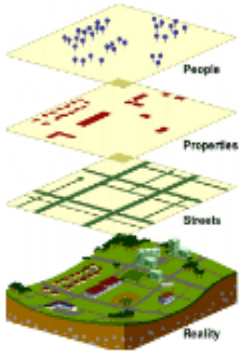
- Partners:
  - ULB - MLG: Machine Learning Group
  - ULB - IGEAT: Institut de Gestion de l'Environnement et d'Aménagement du Territoire
  - RMA - SIC: Signal & Image Centre, Royal Military Academy
  - ULB - SLN: Service Logique et Numérique
- Funding organisation: IRSIB
- Duration: 2007-2009

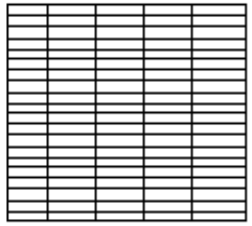
ARMURS

- Project goals:
  - Develop a demonstrator to assist data producers in updating more efficiently their topographic database by using state-of-the-art image processing and statistical analysis techniques
- MLG contribution:
  - Socio-economical and geographical data from sampled communes are being analyzed in order to identify predictive statistical models
  - Different methods will be assessed: rule extraction, decision trees, local learning approaches…

ARMURS

Spatial / Census Data





Spatial Data Mining

Techniques

- Rule extraction
- decision trees
- local learning approaches
- ...

- No change
- Partial change
- Major change

ARMURS

Detection and Analysis of Social Fraud in OASIS Database

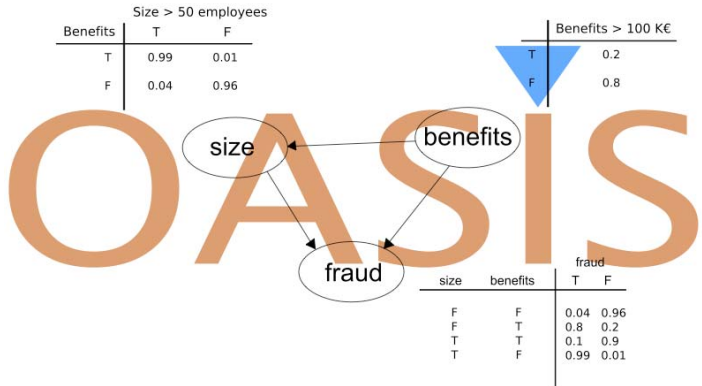
- Partners:
  - UCL - Machine Learning Group (MLG)
  - KUL - Hoger instituut voor de arbeid (HIVA)
  - CNR - Institute of Information Science and Technologies (ISTI)
  - ONSS, SPF SS, SIRS, BCSS, ONEm, SPF Emploi
- Funding organisations: Belgian science policy
- Duration: October 2007 – October 2009

Research content

- OASIS is the data warehouse used by social services to detect social frauds based on report and alarms.
- This project has two main goals :
  - upgrade OASIS by adding data mining tools,
  - better understand the social fraud.
- Mined data cover a large scope from personal data to debts and benefits of societies or unemployment of employees.

Applications

- Build new **alarms** that assist investigators in detecting fraudulent societies
- Outperform present alarms using well suited data mining techniques
- Provide an intuitive interface that outlines the fraudulent societies



Wireless Sensor Networks

- « COMP2SYS - Computational intelligence methods for complex systems » funded by the European Commission (2004 - 2008). Partners: ULB-IRIDIA and ULB-USE.

- « PIMAN - Inspection and Maintenance with Natural Language » funded by Région Bruxelles-Capitale (2007-2008). Partners: Voice Insight, Elsys, VRContext.

Wireless sensor networks

- Wireless sensor modules:
  - Ultra low power devices, with CPU, memory, antenna and sensors
  - Main constraints:
    - **Limited bandwidth for communication**
    - **Limited energy ressources**
- Research axes:
  - Sensor Localization (**manifold learning**)
  - Network organization for efficiently extracting and routing information (**optimization**)
  - In network data processing algorithms (**feature extraction, compression**)

Applications

- Environmental monitoring
- Medical health care
- Battlefield surveillance
- Industrial, structural monitoring (e.g. Breakdown prevention, fault diagnosis)
- Smart homes (Air conditioning control, intelligent lighting)
- …

- More info: *http://www.ulb.ac.be/di/labo*

Feature selection method for calssification problems

- Funding organisation: ULB
- Duration: 2004-2009

Motivations

- More and more problems involve a high number of features
- Discarding irrelevant and redudant features allows to improve classifier performances
- Using information theory is a promising way to determine relevant features in problems involving several thousands of features

Applications

- Biological data analysis (DNA microarray) for classification of tumors
- Analysis of data coming from industrial processes, to determine features involved in
  - Piece defects
  - Process defects
- Data analysis for Customer Relationship Management (CRM)

Molecular Signatures in Human Cancers

- Partners:
  - Microarray Unit (Institut Jules Bordet)
  - Machine Learning Group (Université Libre de Bruxelles)
- Funding organisations:
  - Télévie
  - Les Amis Bordet
  - Fonds Euson, ...
- Duration: 2004-2008

Research content

- Using microarray technology and machine learning methods build a classifier that:
  - can predict patient survival without treatment (prognostic)
  - can predict treatment outcome of patients (prediction)
- Tumors could be classified into subtypes based solely on the difference of expression patterns.

Applications

- Build classifiers based on microarray experiment potentially usable in clinic routines
- Find new drug targets for cancer therapies.
- Improve our understanding of the tumor genesis (implied pathways, tumor subtypes, ...)

Molecular networks in yeast

Integrating experimental and theoretical approaches to decipher the molecular networks of nitrogen utilization in yeast

- Partners:
  - Département de Biologie Moléculaire (B. André)
  - Département de Biologie Moléculaire (J. van Helden)
  - Département de Chimie (M. Kaufman)
- Funding organisations: Communauté Française de Belgique
- Duration: 2004-2009

Research content

- The yeast *Saccharomyces cerevisiae* is the model living organism that is the most extensively used to develop the novel experimental methods of genomics and proteomics.
- We plan *for the first time* to monitor at whole-genome scale the response of yeast cells to supply conditions for a key nutrient of all cellular systems: *nitrogen*.
- Yeast cells are grown under as wide a range of nitrogen supply conditions as possible and the state of their *transcriptome* will be monitored using the *DNA microarray* method.
- Numerical data generated are interpreted by combining statistical, static and dynamic *modelling methods*.
- The project aims to *reverse engineer* the *complete molecular network* of nitrogen metabolism in yeast.

Applications

- Inferring methods developed in this project can be applied:
  - to other genomes, notably the *human genome*.
  - to other domains characterised by numerous features and few observations, notably *web mining*.

Computer-aided medicine

- « Predictive data mining techniques in anaesthesia » FIRST Europe Objectif 1 funded by the Région wallonne and the Fonds Social Européen (2004-2009)
- « TANIA - Decision Support System for Control in Anesthesia » WALEO II project funded by Région Wallonne (2006-2010)
- Partners: University Hospitals of Wallon Region, Mexys SA, Arificial intelligence laboratory in Milano.

Research content

- The goal of the project is to develop **data mining techniques** to analyse the **massive datasets** collected by the "TOOLBOX" software, currently used by the **anestehesists of the Erasme hospital** during the surgical interventions.

- Data concerns the monitored state of the patient (e.g. blood pressure, the rate of heartbeat), the type and concentrations of drugs, the actions of the anesthesist.

Applications

- Build **predictive models** that assist doctors with high level information about the course of the operation.
- Build **visualization application** to help doctors extracting useful information from measured medical data.
- Provide **statistical expertise** to support clinical studies for the certification of new protocols.

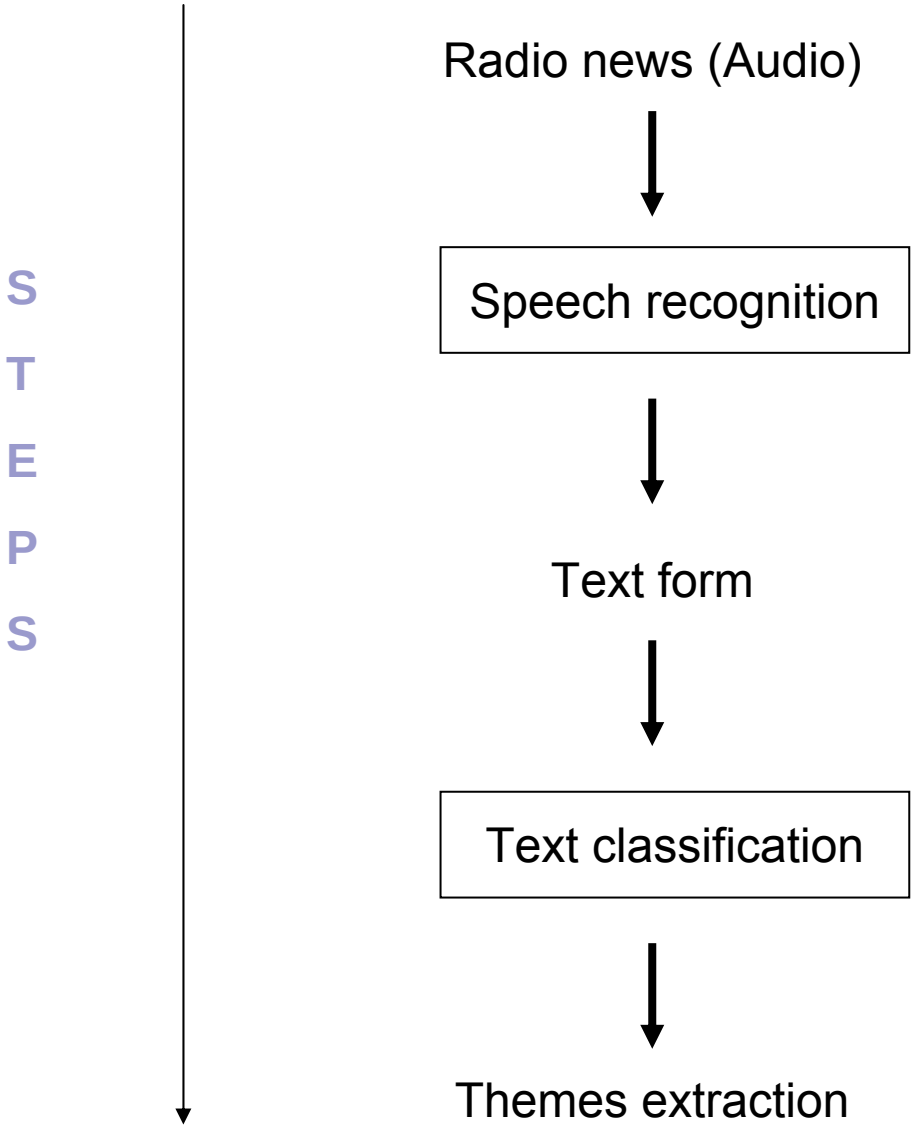
Projet AIDAR

Adressage et **I**ndexation de **D**ocuments multimédias **A**ssistés par des techniques de **R**econnaissance vocale
(Speech Recognition Techniques for Multimedia Document Indexing and Retrieval)

- Partners
  - Voice-Insight s.a. : Speech recongition for database content retrieval
  - asbl TITAN : Digital archiving
- Funding organisation: Région Bruxelles capitale
- Duration: 2004-2006

Researches

- Thematic information extraction from audio and/or text files (Politics, economy, war,…)
- (semi-)automated indexing of audio and text documents (ex : TV news), meta-data generation for content archiving
- Tools :
  - Speech recognition
  - Text classification


S
T
E
P
S
Radio news (Audio)
Speech recognition
Text form
Text classification
Themes extraction

Applications

- Archiving of audio and text documents (Projet ***AIDAR***)
- Filtering and classification of documents for companies :
  - Allow to filter documents received by different services of a company
- Automatic summary of audio/text documents