COMPUTATIONAL INTELLIGENCE GROUP

Pedro Larrañaga
pedro.larranaga@fi.upm.es

Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid
http://cig.fi.upm.es

DIA
Madrid, November 17th, 2010
Outline

1. Members
2. Research Topics
3. Applications
6. Collaborations
27 Members

- **3 Professors**: C. Bielza, P. de Miguel, P. Larrañaga
- **5 Associate Professors**: L. Baumela, J. A. Fernández del Pozo, P. Herrero, J. M. Peña, V. Robles
- **2 Assistant Professors**: J. M. Buenaposada, E. Muñoz
- **4 PostDoc Researchers**: R. Armananzas, A. Latorre, D. Morales, R. Santana
- **12 PhD Students**: F. Baguear, J. Bekios, H. Borchani, A. Férnandez, S. González, L. Guerra, J. Hernando, A. Ibañez, H. Karshenas, P.L. López-Cruz, P. Márquez, D. Vidaurre,
- **1 RTD Management**: P. Flores
Outline

1. Members
2. Research Topics
3. Applications
6. Collaborations
Research Topics

- **Advanced Computation**: parallelize and optimize programs
- **Computer Vision**: efficient algorithms for deformable object registration, extract demographic information and recognize expressions from human faces, image segmentation to reconstruct the cerebral cortex ultrastructure
- **Data Mining**: data streams, multi-dimensional supervised classification, multi-label classification, clustering in high-dimensional spaces, data imputation, feature subset selection using methods as Bayesian networks, regularization, classification by regression
- **Heuristic Optimization**: differential evolution, evolutionary strategies, local search methods, genetic algorithms, estimation of distribution algorithms, particle swarm and gravitational search algorithms
- **Neuroinformatics**: analysis of fMRI or MEG data in order to predict behavioral aspects, morphological cell modeling, neuron classification, spatial synaptic distribution, generate optimized artificial networks with topological features resembling brain networks
Outline

1. Members
2. Research Topics
3. Applications
6. Collaborations
Collaboration agreement with CeSViMa (Centro de Supercomputación y Visualización de Madrid)

Supercomputer Magerit-Intel: 14 eServer BladeCenter HS22, each with 8 processors Intel at 3.2 GHz and 96 GB of shared memory
**Distributed EDAs**

**Voronoi vs. random initializations**

- New procedure to initialize distributed EDAS
- Each population is run on a different processor
Multiple offspring sampling: many techniques to sample new individuals forcing them to compete in the evolutionary process.

Parallel version, with each population in on a different processor.
Optimal use of supercomputers is crucial
Evolutionary hybrid algorithm to optimize process planning in supercomputers
Analysis of human faces to extract demographic information and recognize expressions.
Efficient model-based 3D tracking of deformable objects, in particular of human faces
Electron microscopy image segmentation to reconstruct the cerebral cortex ultrastructure
Selection of human embryos for transfer

Bayesian classifiers to aid to the selection of the most promising embryos that will form the batch to be transferred to the woman’s uterus.
Genotypic predictors of HIV type 1 drug resistance

- Multi-dimensional supervised classification
- 3 groups of drugs: protease inhibitors (PIs), nucleotide reverse transcriptase inhibitors (NRTIs) and nonnucleoside reverse transcriptase inhibitors (NNRTIs)
  \[\Rightarrow 18 \text{ class variables}\]
- Features are mutations
Prognosis of survival in cirrhotic patients

- Feature subset selection by genetic algorithms and estimation of distribution algorithms
Machine learning in bioinformatics

**EVOLUTION**
- Phylogenetic tree construction
- Haplotype inference

**TEXT MINING**
- Word disambiguation
- Knowledge extraction from bibliography and databases (GO, Ensembl, ...)

**PROTEOMICS**
- Protein location prediction
- Protein-protein interaction prediction
- Protein function prediction
- Protein structure prediction

**GENOMICS**
- Gene function prediction
- Operon Id
- TF binding sites Id
- Promoter binding sites Id
- Sequence assembly
- Coding region identification
- Splice site prediction

**FUNCTION PREDICTION**
- Protein function prediction
- Protein structure prediction

**STRUCTURE PREDICTION**
- Protein structure prediction

**MOTIF IDENTIFICATION**
- RNA structure prediction
- SNP's and linkage analysis

**DNA/RNA MICROARRAYS**
- Microarray data pre-processing
- Microarray data analysis
- Microarray based diagnosis & prognosis

**SYSTEMS BIOLOGY**
- Genetic networks
- Signalling networks
- Metabolic pathways

**DIAGNOSIS & PROGNOSIS**
- Mass spectra based diagnosis & prognosis
- Mass spectra pre-processing
- Mass spectra data analysis
- Peptide and protein identification

**METABOLOMICS**
- Chemometric spectral analysis
- Targeted metabolomic profiling and analysis
- Metabolomics simulation programs
- Constraint-based modeling of metabolomic data

**P. Larrañaga**
Computational Intelligence Group (UPM)
Feature subset selection in bioinformatics

A review of feature selection techniques in bioinformatics

P. Larrañaga
Computational Intelligence Group (UPM)
**Data mining: bioinformatics**

**Genomics: DNA microarrays data analysis**

- **Gene selection** approaches in DNA microarray domains
- **Detecting reliable gene interactions** by a hierarchy of Bayesian networks classifiers

P. Larrañaga
Computational Intelligence Group (UPM)
P. Larrañaga

Computational Intelligence Group (UPM)
Autoimmune diseases: Systemic lupus erythematosus

- Select **relevant genes** related with the disease (starting from microarray data)
- Induce **predictive models** (supervised classifiers) to discriminate between **ill and healthy people**
- Obtain knowledge in the form of **genetic networks**
Genomics: genetic (Mendelian) diseases

- Partially supervised approach to dominant and recessive human disease gene prediction
Proteomics: secondary structure prediction

- Bayesian networks as **consensed voting system** in the construction of a multiclassifier for protein secondary structure prediction.
Peakbin selection in mass spectrometry data using a consensus approach with estimation of distribution algorithms
Prediction of citation count of *Bioinformatics* papers

- Predict **citation count** of papers (few, some, many) across the 4-year time horizon
- **Predictive variables**: Tokens found in the abstracts + number of 2-week post-publication periods
- Take into account the journal **section**
High-speed machining

- How to online guarantee a good surface roughness
- Learn a predictive model taking into account cutting parameters and tool variables
Neonatal jaundice

- Decide whether to admit the baby to hospital and the treatment
Primary gastric non-Hodgkin lymphoma

- **Treatment** for its possible cause *Helicobacter pylori* and schedules of chemotherapy and radiotherapy
Heuristic optimization

Optimal row and column ordering

- The optimal table row and column ordering can reveal useful patterns to improve reading and interpretation.
- EDAs and GAs.
Native backbone structure corresponding to the pdb1d2e protein and side chain configuration found by UMDA (EDAs)
Heuristic optimization

EAs for evolution of artificial brain networks

- Evolution of artificial brain networks (simulating the connection topology between different parts of the brain)
- Meet certain existent constraints in the real networks, but they’re optimal wrt other criteria
- Non-dominated solutions learned in each run of the EA (blue points), and absolute non-dominated solutions (filled circles) and original fve32 network (triangle)
Heuristic optimization

Decoding the information contained in different brain waves

- Combination of classifiers with multi-objective optimization algorithms allows to obtain procedures to decode the information contained in different brain waves
- Accuracies using different types of information and all 274 channels (top) vs. occipitoparietal channels (bottom)
Heuristic optimization

Conductance-based compartmental neuron models

- Search for **optimal parameter configurations** and their interactions, reproducing different **neuronal activity** (electrical activity) by using Bayesian networks
Blue Brain Project

- Founded in 2005 by the Brain and Mind Institute of the École Polytechnique in Lausanne (Switzerland), directed by Henry Markram
- An attempt to reverse-engineer the mammalian brain, in order to understand brain function and dysfunction through detailed simulations
- One of the 14 grand challenges for Engineering that will influence science and technology in the 21st century – National Academy of Engineering in 2009, at the request of the NSF

Cajal Blue Brain Project

- UPM and Instituto Cajal (IC) from 2008-2018
- Study the microanatomy and functionality of the column in the somatosensory cortex ⇒ to generate an accurate computer simulation of cortical function
Models and simulation of 3D dendritic tree morphology

Layer III pyramidal neurons from mouse neocortex
Neurons are more varied than cells in any other part of the body

- Morphological features (axonal, dendritic and somatic)
- Interneurons versus pyramidal
A **gardener classification** (not a botanist), coarser and practical

Towards a **consensus** in naming neocortical neurons

- Most neuroscientists agree with terms like pyramidal, nonpyramidal, interneuron, chandelier (clear morphological attributes)
- Other common names like as double bouquet, basket, Martinotti lack a definition and are used inconsistently

Classifications of 320 interneurons given by relevant experts

http://cajalbbp.cesvima.upm.es/gardenerclassification/
Spatial distribution of synapses

- Number and proportion of excitatory (asymmetric) and inhibitory (symmetric) synapses
- Segment them
- Look for spatial patterns (random or specific)
- Density of spines
- How many samples (and their size) to reliably extrapolate
- Distribution of the closest synapses to a given one (important for the neurotransmitter diffusion)
**Neuroinformatics**

**Neurodegenerative diseases: Alzheimer, Parkinson, epilepsy**

- **Discriminate** between AD and controls. Microarray data, fMRI data...
- Define **new scales** to categorize the stage of a Parkinson’s disease patient that consider not only motor features but also cognitive symptoms
- **Predict fully recovery from epilepsy** or not after epilepsy surgery (if drugs are not efficient). Clinical and behavioral test data
“Visualizing” mental activities from brain images and BCI control

- Predict in which direction a subject is covertly attending
- MEG data
Outline

1. Members
2. Research Topics
3. Applications
6. Collaborations


Cuesta, I., Bielza, C., Larrañaga, P., Cuenca-Estrella, M., Laguna, F., Rodríguez-Pardo, D. et al. (2009). Data mining validation of eucast fluconazole breakpoints established by the European committee on antimicrobial susceptibility testing. Antimicrobial Agents and Chemotherapy, 53(7), 2949-2954


B. Calvo, J.A. Lozano, P. Larrañaga (2007). Learning Bayesian classifiers from positive and unlabeled examples. Pattern Recognition Letters, 28(16), 2375-2384


Outline

1. Members
2. Research Topics
3. Applications
6. Collaborations
Projects and Contracts in Progress

Public
- Consolider - Ingenio2010 - CSD2007-00018
- CENIT - Ingenio2010 - 2008 1019
- Cajal Blue Brain
- TIN2008-04528-E
- TIN2008-06815-C02-02
- TIN2010-21289-C02-02
- TIN2010-20900-C04-04
- TRA2008-0148

Private
- Atos Origin
- Banco de Santander
- Panda Security
Outline

1. Members
2. Research Topics
3. Applications
6. Collaborations
Collaborations

- **Atos Origin.** Multilabel classification
- **CSIC - CIEMAT.** Bioinformatics
- **Columbia University.** Neuroscience
- **CSIC. Instituto de Automática Industrial.** Bayesian networks, industry
- **CSIC. Instituto Cajal.** Neuroscience
- **École Polytechnique Fédérale de Lausanne.** Neuroscience
- **Essex University.** Estimation of distribution algorithms
- **Fundación CIEN.** Neuroscience
- **George Mason University.** Neuroscience
- **Hospital de la Paz.** Bioinformatics, neuroscience
- **Hospital de la Santa Creu i Sant Pau.** Neuroscience
- **Hospital Gregorio Marañón.** Decision analysis
- **Kansas University.** Bayesian networks, decision analysis
- **National Research Council of Canada.** Bioinformatics
- **National University of Seoul.** Estimation of distribution algorithms
- **Nijmegen University.** Bayesian networks, neuroscience
- **Panda Security.** Bayesian networks
- **UC San Diego.** Bioinformatics
- **Utrecht University.** Multilabel classification
Pedro Larrañaga
pedro.larranaga@fi.upm.es

Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid
http://cig.fi.upm.es

DIA
Madrid, November 17th, 2010