



DMIP team

Data Mining and Inductive Programming

July 2009



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



DEPARTAMENTO
DE SISTEMAS
INFORMÁTICOS
Y COMPUTACIÓN



Outline

- History
- Members
- Research Areas
- Publications
- Research Projects and Networks
- Committees
- Ph.D. supervision, M.Sc. Teaching, invited talks
- Collaboration with other universities
- Collaboration with companies



History

- Formed in 1997 inside ELP (Extensions of Logic Programming)
- Initially worked on extending *Inductive Logic Programming* to other declarative languages.
- Since then, the team has been acquiring a broader view of the field, exploring different techniques and applications in machine learning and data mining.



Members

- 8 full-time members (six at UPV, two at Universal Nacional de San Luis in Argentina)
- 1 part-time member (also at UPV).
 - M.José Ramírez-Quintana
 - José Hernández-Orallo
 - Cèsar Ferri
 - Vicent Estruch
 - Antonio Bella
 - Ana Funes
 - Arístides Dasso
 - Fernando Martínez-Plumed
 - Josep Alemany
- (Four of them hold a Ph.D.)



Research Areas

- Currently, the research areas are
 - Machine Learning and Data Mining.
 - Knowledge Discovery.
 - ROC analysis, cost-sensitive learning and model evaluation for decision support.
 - Decision trees.
 - Model Calibration.
 - Multi-paradigm Inductive Programming
 - Distance-based methods.
 - Agreement technologies, agents and negotiation.
 - Machine Learning and Artificial Intelligence Evaluation.

Publications

- Books and book chapters:

- E. Hernández, J. Hernández “Programación en C++” Paraninfo, 467 pp, 1993 (1995 2ª Ed.)
- S.Talens, J. Hernández “¡Internet! Redes de Computadores y Sistemas de Información” International Thomson Publishers, 721 pp., 1997 (1998 2ª Ed.)
- E. Hernández, J. Hernández, MC. Juan “C++ Estándar” Thomson Learning Paraninfo, 524 pp., 2001.
- P. Flach, H. Blockeel, C. Ferri, J. Hernández, J. Struyf, “Decision Support for Data Mining: Introduction to ROC analysis and its applications” book chapter in “Data Mining and Decision Support”, Kluwer, 2003.
- J. Hernández, MJ. Ramírez, C. Ferri “Introducción a la Minería de Datos” Prentice Hall, 2004
- J. Hernández, “Knowledge Discovery from Databases” in L.C. Rivero, J.H. Doorn, V.E. Ferraggine (eds.) "Encyclopedia of Database Technologies and Applications", 650 pp. Idea Group, Inc. 2005.
- J. Hernández, “Data warehousing : Multidimensional Data Models and OLAP” in L.C. Rivero, J.H. Doorn, V.E. Ferraggine (eds.) "Encyclopedia of Database Technologies and Applications", 650 pp. Idea Group, Inc. 2005.
- J. Hernández, MJ. Ramírez, “Técnicas de Extracción de Reglas”, in Roque Marín, José Tomás Palma (eds.) “Inteligencia Artificial: Técnicas, Métodos y Aplicaciones”, McGraw-Hill 2008.
- C.Ferri, MJ. Ramírez, “Árboles de Decisión”, in Roque Marín, José Tomás Palma (eds.) “Inteligencia Artificial: Técnicas, Métodos y Aplicaciones”, McGraw-Hill 2008.
- A.Bella, C.Ferri, J. Hernández, MJ. Ramírez, “*Calibration of Machine Learning Models*” in E. Soria, J.D. Martín, R. Magdalena, M.Martínez, A.J. Serrano (eds.) " Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques ", IGI Global 2009.

Publications

- **Journals and Conferences:**
 - More than 20 publications in journals or volumes listed in SCI (Science Citation Index), or SCI Expanded), such as “Artificial Intelligence” (Elsevier), “Journal of Intelligent Systems” (John Wiley), “Pattern Recognition Letters” (Elsevier), “AI Communication” (IOS Press)...
 - About 10 publications in journals which are not included in SCI.
 - More than 70 conferences (almost all of them international).
 - More than 500 citations to articles written by members of the group. Most-cited paper: C. Ferri, P. Flach and J. Hernández Orallo “Learning decision trees using the area under the ROC curve”, ICML’02 (Core A+): 137 citations.



Research Projects and Networks

- **Coordinated by the group (selection)**

- MEDIM: Minería de datos estructurados basados en distancias y métricas (GV04B/477)
- META-MIDAS: Mètodes d'acomodació per a una mineria de dades sinèrgica (GV04A/389)
- OMID: Optimización de Modelos de Minería de Datos (GV 06/301)
- ANYNT (Anytime Intelligence): Min. Ciencia e Innovación, Convocatoria EXPLORA (TIN2009-06078)

- **Participations (selection)**

- 4 national projects
- 1 national “excellence” projects (CONSOLIDER INGENIO): “Agreement Technologies”.
- European networks: ILPNET2, KdUbiq
- EU funding: “ICT for EU-India Cross-Cultural Dissemination” and “Alfa – LERNet”).
- Several “integrated actions” (with Germany, Austria and Italy).
- Part of the “European Research Consortium for Informatics and Mathematics” (ERCIM).

Committees

- Members of the group have participated in many programme committees and organising committees:
 - Programme Committees: STAIRS'2004, ICML'2004, ROCAI'2004, AAIP'2005, ICML'2005, ROCML'2005, CAEPIA'2005, TAMIDA'2005, NNN'2005, ECAI'2006, ECML'2006/PKDD'2006, ROCML'2006, WTDIA'2006, CTDIA'2006, IDEAL'2006, ECML'2007/PKDD'2007, CAEPIA'2007, ECAI'2008, ICML'2008, Iberamia'2008, AI'08, KDD'09, CAEPIA'09, AGI-09, AAIP'09.
 - Organising Committees: WFLP'2000, RDP'2003, WFLP'2003, STAIRS'2004 (ECAI'2004), ROCAI'2004, WWV'2005, PLID2008, LOPSTR2008, SAS2008, PPDP2008.
 - Journal reviewers for: "Automated Software Engineering", Kluwer Academic Publishers, "Machine Learning Journal", Kluwer Academic Publishers, "Theory and Practice of Logic Programming", Cambridge University, "Journal of Machine Learning Research", MIT, "Computación y Sistemas", "ACM Computing Reviews", "Inteligencia artificial, Revista Iberoamericana de IA", Neurocomputing, Data Mining and Knowledge Discovery and Journal of Artificial Intelligence Research. *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Medical physics.
 - Conference reviewers for: AGP'2000, WFLP'2001, PPDP'2001, SAC'2002, WFLP'2002, LPAR'2002, RTA'2003, IJCAI'2003, JISBD'2002, PROLE'2002, CAEPIA'2003, RTA'2003, IJCAI'2003, AGP'2003, IEA/AIE 2004, JISBD'2004, CAISE 2005, IJCAI'2005, ECML'2005, IJCAI'2006, AI'2006, FLOPS2007, SAS2009.



Ph.D supervision, M.Sc. Teaching and invited talks

- 3 Ph.D. students completed their Ph.D. inside the team.
- 4 students are currently performing their Ph.D. under the supervision of a team's member.
- Several invited lectures at national and foreign institutions:
 - (Vitus Bering (Denmark), University of Bristol (UK), University of El Salvador (El Salvador), University of Fco. Sta. Maria (Chile), Universidad de La Laguna (Spain), Universidad de Castilla La Mancha (Spain), Instituto Tecnológico de Chihuahua (Mexico), Università di Roma Tre (Italy), TU Wien (Austria), Universidad de Sevilla (Spain), Università degli Studi di Udine (Italy), Deutsche Telekom / FachHochschule Kommunikation Leipzig (Germany), among others (Bolivia, Pamplona, ..).



Collaboration with other universities

- Since its inception, the team has established research contacts and collaborations with some of the most important groups in our research area. Because of this, we have maintained or maintain collaboration with the following entities:
- Fraunhofer Institut Autonome Intelligente Systeme (Germany): in collaboration with Thomas Gärtner.
- European Research Consortium for Informatics and Mathematics (Europe)
- Institut d'Investigació en Intel·ligència Artificial. CSIC (Spain)
- University of Bristol (UK): Peter Flach's group.
- Monash University (Australia): David Dowe's group.
- Università di Siena (Italia): Moreno Falaschi's group.
- Universidad EAFIT de Medellin (Colombia)
- Universidad Nacional de San Luis (Argentina)
- FachHochschule Kommunikation Leipzig (Germany)



Collaboration with companies

- We have had more than 10 technology transfer contracts with several companies, from several sectors.
- Some of these companies are:
 - Dimensión Informática (software engineering).
 - Hospital Clínico Universitario de Valencia (healthcare).
 - Mercadona (retailing).
 - ADD (software engineering).
 - Deutsche Telekom (data mining applied to the building industry)