A Report of Activities at the WIC-Spain Research Centre

Pilar Herrero, María S. Pérez, Ernestina Menasalvas, Javier Segovia
Universidad Politécnica de Madrid, Spain

1 What is the WIC-Spain Research Centre?

The WIC(Web Intelligence Consortium)-Spain Research Centre is an interdisciplinary research initiative which brings together people with the aim of covering all the aspects related to the discovery of hidden knowledge in big volumes of data as long as exploring the practical impacts of advanced Information Technology (IT) and Artificial Intelligence (AI) on this volume of information

The group consist of seven researchers and five Ph.D. students led by Dr. Javier Segovia and coordinated by Dr. Ernestina Menasalvas. We are funded from a variety of sources including the Spanish Ministry of Science and Technology as well as national and international industry.

2 Who are the members of the WIC-Spain Research Centre?

The group is led by Dr. Javier Segovia and coordinated by Dr. Ernestina Menasalvas.

The other current members of the research group are:

- **Researchers:**
    - Dr. Santiago Eibe
    - Dr. Pilar Herrero
    - Dr. Óscar Marbán
    - Dr. José M. Peña
    - Dr. María S. Pérez
    - Dr. Víctor Robles
    - Dr. Marta del Socorro

- **Phd Students:**
    - Óscar Cubo
    - Pedro González
    - Vanessa Herves
    - Esther Hochsztain
    - Alberto Sánchez

- **Secretary:**
    - Coro Pérez

3 What is the origin of the WIC-Spain Research Centre?

Our research group has been set up in 2002 in the Facultad de Informática at the Universidad Politécnica de Madrid (Spain), mixing Dr. Javier's background (in areas such as Artificial Intelligence, System Modeling, Virtual Environments and Data Mining) with the background of the rest of the people involved in the group (Data Mining, Data Bases, Data Warehousing, Heuristic Optimization, Intelligent Agents and Multi-Agents Systems and Grid Computing).

4 What are the WIC-Spain's research lines?

Currently, we are focused on the development and application of Data Mining to the following research lines:

4.1 Social Simulation

There is growing interest in using computer simulation to explore issues in most parts of the social sciences, including sociology, political science, economics, anthropology, geography, archaeology and linguistics. It can also be the inspiration for new, process-oriented theories of society.

Based mainly on Official Data Bases - such as Population and Housing Censuses or Family Expenditure Surveys (IPC) - this research line uses techniques of Artificial Life (AL) to develop models of nonhuman life-forms, Cellular Automatas (CA) to model on a grid with local interactions between cells, Distributed Artificial Intelligence to introduce intelligent interactions, Genetic Algorithms to learn through evolution and selection, Microsimulation to model, year by year, taxes, benefits and changing individual attributes, Neural Networks to learn and System Dynamics to introduce models of equations of macro-level differential and difference.

We have already received national and international funding for working on this research line, having developed some applications for micromarketing that makes possible, for example, to analyse the life cycle model of a customer, demonstrating his tendency of purchasing.

IEEE
COMPUTER
SOCIETY

Currently, as a result of this research work, ICC+ (©2003) have been obtained in collaboration with private funding (Boole@).

Nowadays, we are working on an extension of the method/technologies to European Level, studying the general evolution of the society (geographical & behavioral).

4.2 Data Quality

The fundamental goal of this research line is to improve the quality of customers data, by using appropriated processing techniques - such as Cleaning, Normalization or DeDuplication - or by trying new methods -such as Neural Networks or Genetic Programming.

At present, we have been working on projects, related to this research area, with national and international funding.

4.3 Process for KD&DM

The WIC(Web Intelligence Consortium)-Spain Research Centre has established a research line to explore some of the problems found in the application of CRISP-DM, broaching problems such as the lack of techniques for business modelling, the deployment of data models (mining models), software engineering for these applications, project estimation or project organisation, and looking for the inspiration source in important research areas such as software engineering, knowledge engineering, DM process & methodologies and CRM methodologies [1]. The group is currently applying its results in a project funded by a Malaysian company.

4.4 BioMining

Two projects are currently under development. One of them has as main goal the Protein Secondary Structure Prediction (PSSP). Different approaches and techniques can be used depending on the data structure: a new method based on Bayesian networks and multi-classifier [4]. We have already obtained excellent results with the multi-classifier approach. The second project is being performed in cooperation with a local Hospital in Madrid.

4.5 WebMining

The main objective of this research line is to calculate typologies of navigators according to factors such as navigation, behavior, business goals and visited pages, estimating the value of ongoing sessions and integrating the web mining with some business goals.

We are working in the research line in cooperation with the University of ORT (Urugay) and the University del Valle of Cali (Colombia).

4.6 Modelling Systems for Engineering

Starting for an international collaboration established in 1992 with the University of Tennessee College of Engineering, Knoxville , Tennessee, the Oak Ridge National Laboratory and the WIC-Spain research group work together on neural networks and genetic programming techniques.

4.7 Data Grid

Due to the necessity of increasing the computational resolution of complex applications, grid computing is becoming a new emergent research area. The Grid Computing challenge is the complete integration of heterogeneous computing systems and data resources with the aim of providing a global computing space with an efficient access to data [3]. We are working in this area in cooperation with the University of Nottingham (in UK) and the Decision System Group (Harvard/MIT).

5 The WIC-Spain's Activities

The WIC-Spain Research Centre is continuously organizing scientific activities. In this way, we have already been involved in the organization of AWIC'03 [2] and AWIC'04, and, currently, we are working in the organization of the First International Workshop on Grid Computing and its Application to Data Analysis (http://laurel.datsi.fi.upm.es/GADA04/) to be held in Cyprus and the International Symposium on Intelligent Data Analysis (IDA'05), which will be held in Spain in 2005.

References

[1] C. Fernández, E. Menasalvas, A. Wasilewska, and J.M. Peña. The lattice structure of KDD process: Mathematical expresion of the KDD process. *Journal of International Computer Science*, 2001.

[2] E. Menasalvas, J. Segovia, and P.S. Szczepaniak. Web intelligence, first international atlantic web intelligence conference, AWIC 2003. *Proceedings Springer*, 2003.

[3] M.S. Pérez, J. Carretero, F. García, J.M. Peña, and V. Robles. MAPFS-Grid: A flexible architecture for data-intensive grid applications. *Grid Computing*, pages 111–118, 2004.

[4] V. Robles, P. Larrañaga, J.M. Peña, E. Menasalvas, and M.S. Pérez. Interval estimation naïve bayes. *Intelligent Data Analysis 2003*, pages 143–154, 2003.

2