Adaptive Business Intelligence

A Technology Curricular Unit proposal for the MAP-I Phd Program

A – Programmatic Component

1. Motivation <mark>(tema, justificação e enquadramento: áreas científicas cobertas, disciplinas similares em projectos de ensino comparáveis internacionalmente)</mark>

Nowadays, business organizations are increasingly moving towards decision-making processes that are based on information. **Business Intelligence (BI)** is an umbrella term that includes methodologies, architectures, tools, applications and technologies to enhance managerial decision making [1]. The goal of BI is to: access data from multiple sources, transform these data into information and then into knowledge.

Very recently, a new trend emerged in the marketplace called **Adaptive Business Intelligence (ABI)** [2]. Besides transforming data into knowledge, ABI also includes the decision-making process. BI systems often include elements of databases, data warehouses and data Mining [1], while ABI systems also encompass forecasting [3] and optimization [4], in order to enhance adaptability. In effect, adaptability is a vital component of any intelligent system and this issue is expected to gain popularity in the next years. The final ABI goal is to use computer systems that can adapt to changes in the environment, solving complex real-world problems with multiple objectives, in order to aid business managers to make better decisions, increasing efficiency, productivity and competitiveness.

Although being a recent field, the topics covered by ABI (e.g. data mining, forecasting, modern optimization) have a large research community, with several prestigious international scientific journals (e.g. IEEE Trans. Evolutionary Computation, Machine Learning, IEEE Trans. Neural Networks, Data Mining Knowledge Discovery, Decision Support Systems) and conferences (e.g. ACM KDD, ACM CIKM, ACM ICIS, IEEE ICDM, IEEE CEC, IEEE IJCNN) available.

There are also several international examples of Computer Science PhD programs that include ABI topics, such as:
- Carnegie Mellon University (CMU), USA:
  - **Ph.D. Program in Computer Science** (optimization);
  - **Ph.D. Program in Computation, Organizations and Society** (optimization, knowledge management, machine learning);
  - **Ph.D. Program in Machine Learning** and **Joint Ph.D Program in Statistics & Machine Learning** (machine Learning, data mining, database management systems);
- Berkeley University of California, USA:
  - **Ph.D. in Computer Science**, specialization ins Communication, Computation and Statistics (forecasting, machine learning, decision-making);
- University of Texas at Austin, USA:
  - **Ph.D. in Computer Science** (database systems, machine learning, data mining, neural networks);
- Auburn University, USA:
  - **Ph.D. in Computer Science and Software Engineering** (artificial

intelligence and genetic algorithms, database and applications)

2 Objectives and Learning Outcomes

To learn about the basic ABI concepts, including: characteristics of complex business problems, BI and ABI, data mining, prediction, optimization and adaptability;

To master the state of the art methods and models (e.g. Decision Trees, Neural Networks, Support Vector Machines, Learning Classifier Systems, Evolutionary Algoritms), methodologies (e.g. CRISP-DM), and tools (e.g. R, WEKA, SAS, Evolution Machine, SCS-C);

To apply ABI in real-world applications (e.g. Finance, Economy, Marketing); and

To perform an essay over an advanced research ABI topic (e.g. mining complex data; forecasting using sophisticated machine learning methods; multi-objective evolutionary optimization; adaptive decision-making).

3 Detailed Program

1 - Introductory ABI concepts: characteristics of complex business problems, BI and ABI, ABI case studies, data mining, prediction, optimization and adaptability.
2 - Modern Learning and optimization methods: decision trees, neural networks, support vector machines, nearest neighbors, evolutionary algorithms, learning classifier systems.
3 - Data mining: CRISP-DM, knowledge understanding, data selection, data preprocessing; data mining; data interpretation and knowledge maintenance.
4 - Forecasting: univariate and multivariate forecasting, time series, conventional forecasting methods.
5 - Exploration of ABI tools (e.g. R, WEKA, SAS, Evolution Machine, SCS-C) when applied to real-world problems (e.g. Finance, Economy, Marketing).
6 - Perform a research essay over an advanced ABI topic.

4 Teaching Methodology and Evaluation

Four teaching methodologies will be applied:
1 Lecture exposition of key ABI issues.
2 Active learning (e.g. think-pair-share, in-class teams [5]).
3 Case-based learining.
4 Project based learning;

Evaluation will include three elements:
      A - review of key ABI research articles (25%);
      B - an ABI project that describes the application of ABI tools to real-world datasets (50%); and
      C - research essay (25%).

5 Bibliography (up to 10 references)

[1] E. Turban, R. Shard, J. Aronson and D. King, Business Intelligence – A

Managerial Approach, Pearson Prentice-Hall, New Jersey, USA, 2008.

[2] Z. Michalewicz, M. Schmidt, M. Michalewicz and C. Chiriac, Adaptive Business Intelligence, Springer-Verlag, Leipzig, Germany, 2007.

[3] S. Makridakis, S. Wellwright and R. Hyndman, Forecasting: Methods and Applications, John Wiley & Sons, New York, USA, 1998.

[4] S. Ruhul, M. Masoud and X. Yao (Eds.), Evolutionary Optimization, Springer-Verlag, 2002.

[5] D. Johnson, R. Johnson and K. Smith, Active Learning: Cooperation in the College Classroom, 2nd edition, Edina, Interaction Book Company, 1998.

B Lecture Team

1. Summary (Uma apresentação sumária da equipa docente proponente e suas valências, fornecendo explicitamente evidência de investigação corrente e activa na área que a Unidade Curricular proposta entende cobrir.)

The team is willing to write didactic texts related to this unit.

If necessary and funds are available, one of the team members can visit CMU for the curricular unit accreditation.

2. Coordinator

Manuel Filipe Santos (MFS)

3. CVs (CV resumidos de todos os membros da equipa, referindo, em particular, publicações, projectos de investigação financiados e orientações de doutoramento nos últimos 5 anos.)

3.1 Manuel Filipe Santos

Biography: Manuel Filipe Santos received his Ph.D. in Computer Science (Artificial Intelligence) from the University of Minho (UMinho), Portugal, in 2000. He is auxiliary professor at the Department of Information Systems, UMinho, teaching undergraduate and graduate classes of Business Intelligence and Intelligent Data Analysis. He his also researcher at the Business Intelligence Group (big.dsi.uminho.pt) of the R&D Algoritmi Centre, with the current research interests:

- Business Intelligence and Decision Support Systems;
- Data Mining and Machine Learning (Learning Classifier Systems);
- Grid Data Mining.

Relevant publications in the last 5 years:

His most significant publications for the field in the last 5 years are:

[1] Á. Silva, **P. Cortez**, M.F. Santos, L. Gomes and J. Neves. Rating organ failure via adverse events using data mining in the intensive care unit. In *Artificial Intelligence in Medicine*, Elsevier, *In Press* (*ISI impact factor 1.882*).

[2] H. Quintela, M. F. Santos and **P. Cortez**. Real-Time Intelligent Decision Support System for Bridges Structures Behavior Prediction. In J. Neves, M. F.

Santos and J. Machado (Eds.), Progress In Artificial Intelligence, **13th EPIA Portuguese Conference on Artificial Intelligence**, Lecture Notes in Computer Science 4874, pp. 124-132, Guimarães, Portugal, December, 2007. Springer, ISBN 978-3-540-77000-8 (*ISI proceedings*).

[3] M.F. Santos, P. Cortez, J. Pereira and H. Quintela. Corporate Bankruptcy Prediction using Data Mining Techniques1. In A. Zanasi, C. Brebbia and N. Ebecken (Eds.), **Data Mining VII - Data, Text and Web Mining, and their Business Applications**, WIT Transactions of Information and Communication Technologies, vol. 37, pp. 349-357, 2006. WIT Press, UK, ISBN:1-84564-178-7, ISSN:1743-4463 (*ISI proceedings*).

[4] Á. Silva, **P. Cortez**, M.F. Santos, L. Gomes and J. Neves. Mortality assessment in intensive care units via adverse events using artificial neural networks. In *Artificial Intelligence in Medicine*, Elsevier, 36 (3): 223-234, 2006 (*ISI impact factor 1.882*).

[5] Neves, J., M.Santos, Machado, J., (Eds.), Progress in Artificial Intelligence, **13th EPIA Portuguese Conference on Artificial Intelligence**, Lecture Notes in Computer Science 4874, Guimarães, Portugal, December, 2007. Springer, ISBN 978-3-540-77000-8.

[6] Neves, J., Santos, M.F., Machado, J., (Eds) New Trends in Artificial Intelligence, December 2007, APPIA, Portugal.

[7] Abelha A., Analide C., Machado J., Neves J., Santos M., Novais P., AMBIENT INTELLIGENCE AND SIMULATION IN HEALTH CARE VIRTUAL SCENARIOS, in Proceedings of 8th IFIP Conferences on Virtual Enterprises, Springer, Germany, 2007.

[8] Gago, P. Silva, A., Santos, M., Adaptive Decision Support for Intensive Care, Proceedings of 13th Portuguese Conference on Artificial Intelligence, Springer, Germany, 2007.

[9] Quintela, H, Santos, M., Cortez, P., Real-Time Intelligent Decision Support System for Bridges StructuresBehavior Prediction, Proceedings of 13th Portuguese Conference on Artificial Intelligence, Springer, Germany, 2007.

Participation in R&D projects in the last 5 years:

He participated in various R&D projects, being Principal Investigator of 2 FCT financed projects:

INTCARE – Intensive Care Intelligent Decision Support System, PTDC/EIA/72819/2006;

GridClass – Learning Classifiers for Grid Data Mining, GRID/GRI/81736/2006.

Supervision of Graduate Students:

Supervised 10 MSc theses and 2 PhD theses.

Other relevant topics of his CV:
- **Co-organized** the EPIA 2007 – 13th Portuguese Conference on Artificial Intelligence.

3.2 Paulo Cortez

Biography: Paulo Cortez (www.dsi.uminho.pt/~pcortez) received his Ph.D. in Computer Science from the University of Minho (UMinho), Portugal, in 2002. He is lecturer at the Department of Information Systems, UMinho, teaching undergraduate and graduate classes of Computer Programming and Business Intelligence. He his also researcher at the Business Intelligence Group of the R&D Algoritmi Centre, with the current research interests:

- Business Intelligence and Decision Support Systems;
- Data Mining and Machine Learning;
- Neural Networks and Evolutionary Computation;
- Forecasting.

Relevant publications in the last 5 years:

He is co-author of more than fifty publications in international conferences and journals (e.g. published by IEEE, Elsevier or Springer). His relevant publications in the last 5 years are:

[10]   Á. Silva, **P. Cortez**, M.F. Santos, L. Gomes and J. Neves. Rating organ failure via adverse events using data mining in the intensive care unit. In *Artificial Intelligence in Medicine*, Elsevier, *In Press* (*ISI impact factor 1.882*).

[11]   **P. Cortez**, RMiner: Data Mining with Neural Networks and Support Vector Machines using R, book chapter of *Introduction to Advanced Scientific Softwares and Toolboxes*, International Association for Engineering (IAE), 15 pages, *In Press*.

[12]   P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In **Proceedings of FUture BUsiness TEChnology Conference (FUBUTEC 2008)**, Porto, April, 2008 (*ISI proceedings*).

[13]   H. Quintela, M. F. Santos and **P. Cortez**. Real-Time Intelligent Decision Support System for Bridges Structures Behavior Prediction. In J. Neves, M. F. Santos and J. Machado (Eds.), Progress In Artificial Intelligence, **13th EPIA Portuguese Conference on Artificial Intelligence**, Lecture Notes in Computer Science 4874, pp. 124-132, Guimarães, Portugal, December, 2007. Springer, ISBN 978-3-540-77000-8 (*ISI proceedings*).

[14]   M. Rocha, **P. Cortez** and J. Neves. Evolution of Neural Networks for Classification and Regression. In *Neurocomputing*, Elsevier, 70 (16-18):2809-2816, October, 2007 (*ISI impact factor 0.790*).

[15]   **P. Cortez**, M. Rio, P. Sousa and M. Rocha. Topology Aware Internet Traffic Forecasting using Neural Networks. In J. de Sá et al. (Eds.), **Artificial Neural Networks - ICANN, 17th International Conference**, Lecture Notes in Computer Science 4669, pp. 445-454, Porto, Portugal, September, 2007. Springer, ISBN: 978-3-540-74693-5.

[16]   D. Duque, H. Santos and P. Cortez. Prediction of Abnormal Behaviors for Intelligent Video Surveillance Systems. In **Proceedings of the 2007 IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)**, pp. 362-367, Honolulu, USA, April, 2007. IEEE, ISBN: 1-4244-0698-6 (*ISI proceedings*).

[17]   **P. Cortez**, M. Rio, M. Rocha and P. Sousa. Internet Traffic Forecasting using Neural Networks. In **Proc. of the 2006 IEEE Int. Joint Conference on Neural Networks**, Vancouver, Canada, pp. 4942-4949, July, 2006, IEEE (*ISI proceedings*).

[18]   M.F. Santos, P. Cortez, J. Pereira and H. Quintela. Corporate Bankruptcy Prediction using Data Mining Techniques1. In A. Zanasi, C. Brebbia and N. Ebecken (Eds.), **Data Mining VII - Data, Text and Web Mining, and their Business Applications**, WIT Transactions of Information and Communication Technologies, vol. 37, pp. 349-357, 2006. WIT Press, UK, ISBN:1-84564-178-7, ISSN:1743-4463 (*ISI proceedings*).

[19]   Á. Silva, **P. Cortez**, M.F. Santos, L. Gomes and J. Neves. Mortality assessment in intensive care units via adverse events using artificial neural networks. In *Artificial Intelligence in Medicine*, Elsevier, 36 (3): 223-234, 2006 (*ISI impact factor 1.882*).

[20]   **P. Cortez**, M. Rocha and J. Neves. Evolving Time Series Forecasting ARMA Models, In *Journal of Heuristics*, Springer, 10 (4): 415-429, 2004 (*ISI impact factor 1.113*).

Participation in R&D projects in the last 5 years:

He participated in 7 R&D projects, being Principal Investigator of 2 projects, namely:

- **Principal Investigator (PI)** of the project **PTDC/EIA/64541/2006 - SPAM Telescope Miner: worldwide unsolicited email detection using data mining techniques**, financed by FCT, from January 2008 to December 2010. Budget: 70000 euros.
- **Principal Investigator (PI)** of the Portuguese team in the project **B-53/05 - Internet Congestion Control Using Neural Networks**, financed by the *Portuguese Association of Rectors (CRUP)/British Council*, from April 2005 to April 2006. This project involved UMINHO and University College London.

Supervision of Graduate Students:

Supervised 6 MSc thesis and currently he is finishing a co-supervision of 1 PhD thesis: Duarte Duque, Previsão e Identificação de Eventos de Quebra de Segurança em Vídeo-Vigilância, co-supervision with Prof. Henrique Santos, Ph.D. in Information Systems and Technologies, University of Minho (started in 2005, FCT grant).

Other relevant topics of his CV:

- **Associate Editor** of the **Neural Processing Letters** journal (Springer, ISI).
- **Reviewer** of several ISI journals (e.g. Neural Processing Letters, Artificial Intelligence in Medicine) and conferences (e.g. IJCNN, DMin, ICDM, PPKDD).
- **Co-organized** the Business Intelligence BI 2007 workshop of EPIA and Neural Networks in Biomedical Engineering and Bioinformatics workshop of ICANN07;
- **Invited lecturer** in the Int. Summer School of Neural Networks in Classification, Regression and Data Mining (2003-06;2008).
- Acted as **external examiner** of 2 MSc and 2 PhD thesis.
- **Author** of the open source RMiner library, which facilitates the use of Data Mining applications in R.

3.3 Rui Camacho
Biography: Rui Camacho received his Ph.D. in Electrical Engineering and Computers from the University of Porto (UP), Portugal, in 2000. He is auxiliary

professor at the Informatics Engineering Department, UP, teaching undergraduate and graduate classes of Machine Learning and Data Mining. He his also researcher at the Laboratory of Artificial Intelligence and Decision Support (LIAAD), with the current research interests:

- Inductive Logic Programming;
- Data Mining and Machine Learning;
- Data Mining;
- Applications of Bioinformatics.

Relevant publications in the last 5 years:

His most significant publications for the field in the last 5 years are:

[21]   editor with, Joao Gama, Luis Torgo, Alipio Jorge and Pavel Brazdil, of the ECML 2005 Proceedings.

[22]   editor with, Joao Gama, Luis Torgo, Alipio Jorge and Pavel Brazdil, of the PKDD 2005 Proceedings.

[23]   editor, with Ahsiwn Srinivasan and Ross King, of the proceedings of the ILP 2004.

[24]   Nuno A. Fonseca, Fernando Silva, Rui Camacho and Ashwin Srinivasan, *Parallel ILP for Distributed-Memory Architecture,* in Machine Learning journal (to appear in 2008).

[25]   Ruy Ramos, Rui Camacho, *A step up with the HARVARD system: the HARVARD-g system*, 1st Iberian GRID Infrastructure Conference Proceedings (IBERGRID 2007), pp 405-408, Santiago de Compostela, Espanha, 2007.

[26]   Ruy Ramos and Rui Camacho, *Distributed Generative Data Mining*, 7th Industrial Conference on Data Mining (ICDM~2007), Springer-Verlag LNAI 4597, pp 307-317, Leipzig/Germany, July 14-18, 2007.

Participation in R&D projects in the last 5 years:

He participated in various R&D projects, being Principal Investigator of 1 project, namely:

ILP-Web-Service: An Inductive Logic Programming based Web service

Supervision of Graduate Students:

Supervised 2 MSc theses and 1 PhD theses. Currently supervises 3 PhD students and 3 post-doc students.

Other relevant topics of his CV:

- **Co-organized** the ILP 2004 – International Conference on Inductive Logic Programming.
- **Co-organized** the ECML/PKDD 2005 – Europena Conference on Machine Learning and the European Conference on Principles and Practice of Knowledge Discovery in Databases.
- Was guest editor of the Machine Learning journal Vol. 64, N. 1/2/3, 2006.