Adaptive Business Intelligence (ABI)

Technology Curricular Unit (UTC) proposal for the MAP-I PhD Program

A – Programmatic Component

1. Motivation

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 encompass two additional modules: **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 (i.e. data mining, forecasting, modern optimization and adaptive systems) have a large research community, with several prestigious international scientific journals (e.g. Data Mining Knowledge Discovery, Decision Support Systems, Machine Learning, IEEE Trans. Neural Networks, International Journal of Forecasting, IEEE Trans. Evolutionary Computation, Journal of Heuristics, Applied Soft Computing) and conferences (e.g. ACM KDD, ACM CIKM, ACM ICIS, IEEE ICDM, IEEE IJCNN, IEEE CEC) 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** (machine learning, optimization);
  - **Ph.D. Program in Computation, Organizations and Society (**advanced artificial intelligence, machine learning);
  - **Ph.D. Program in Machine Learning** (data mining, database management systems, machine Learning, optimization);
- Standford University, USA:
  - **Ph.D. in Computer Science** (databases, data mining, machine learning, modern optimization)
- Berkeley University of California, USA:
  - **Ph.D. in Computer Science**, specialization ins Communication, Computation and Statistics (database management systems, data mining, decision support);
- University of Texas at Austin, USA:
  - **Ph.D. in Computer Science** (database systems, data mining, machine

learning, neural networks, reinforcement learning);

- Cornel University, USA:
  - **Ph.D. Program in Information Science** (advanced artificial intelligence, advanced databases, machine learning).

The proposed ABI unit had already **four** previous **MAP-I editions**:

- **2012/13** lective year, at University of Aveiro (http://www.map.edu.pt/i/2012/Courses);
- **2011/12** lective year, at University of Porto (http://www.map.edu.pt/i/2011/Courses);
- **2010/11** lective year, at University of Minho (http://www.map.edu.pt/i/2010/Courses);
- **2008/09** lective year, at University of Porto (http://www.map.edu.pt/i/2008/courses).

The **assessment made by the students** on the previous editions **encourages further editions**. An anonymous questionnaire was launched in the e-learning system and the student's average responses were:

- Question: "This teaching unit is useful for the PhD program". Average responses: year of 2012/13 - 87% (highly agree); over all lective years - 83% (highly agree).
- Question: "The teaching materials and resources provided by the teacher were adequate". Average responses: year of 2012/13 - 80% (highly agree); over all lective years - 83% (highly agree).
- Question: "Positive stimulus for an active student participation and discussion in class?" Average responses: year of 2012/13: 87% (highly agree); over all lective years - 89% (completely agree).
- Question: "Positive evaluation of the teachers global performance in this teaching unit". Average responses: year of 2012/13: 92% (completely agree); over all lective years - 87% (highly agree).

2 Objectives and Learning Outcomes

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

To master the state of the art of ABI methods and models and tools;

To perform a review essay over an advanced research ABI topic;

To apply ABI in real-world applications.

3 Detailed Program

1 - Introductory ABI concepts: characteristics of complex business problems, BI and ABI, data mining, prediction, optimization and adaptability, state of the art.
2 – Using prediction and optimization to build adaptive systems: application of data mining models and techniques in ABI (e.g. decision trees, neural networks, support vector machine, learning classifier systems, hierarchical and relational clustering, inductive logic programming), application of optimization techniques in ABI (e.g., heuristic search, hill-climbing, tabu-search, evolutionary computation).
3 - Conducting ABI projects and case studies: CRISP-DM, ABI applied to real-world problems (e.g. Finance, Economy, Marketing).
4 - Exploration of ABI tools: DM and optimization tools (e.g. R, WEKA, SAS Enterprise Miner, Rapidminer, Evolution Machine, SCS-C, Aleph, Moss), BI tools (e.g., SAS, MS SQL Server).

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 learning.
4 Project based learning.

Evaluation will include three elements:

A - review of ABI research article(s) (20%); and
B - an ABI project that describes the application of ABI tools to real-world datasets (80%);

5 Bibliography

Cited references:

[1] E. Turban, R. Sharda, J. Aronson and D. King, Business Intelligence – A Managerial Approach, Pearson Prentice-Hall, New Jersey, USA, 2010.
[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] Z. Michalewicz and D. Fogel, How to solve it: modern heuristics. Springer, 2004.
[5] D. Johnson, R. Johnson and K. Smith, Active Learning: Cooperation in the College Classroom, 2nd edition, Edina, Interaction Book Company, 1998.

Additional references:

[6] S. Luke, S. Essentials of metaheuristics. George Mason University. Free access: http://cs.gmu.edu/~sean/book/metaheuristics/, 2009.
[7] Michalewicz, Z., Schmidt, M., Michalewicz, M. and Chiriac, C. (2005). Case study: an intelligent decision support system. In IEEE Intelligent Systems, 20(4):44-49.
[8] Michalewicz, Z., Schmidt, M., Michalewicz, M. and Chiriac, C. (2007). Adaptive Business Intelligence: Three Case Studies, In Evolutionary Computation

in Dynamic and Uncertain Environments, pp. 179-196, Springer.
[9] E. Turban, R. Sharda and D. Delen. Decision Support and Business Intelligence Systems, Prentice Hall, 2011.
[10] P. Cortez, Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool, In P. Perner (Ed.), Advances in Data Mining, Proceedings of 10th Industrial Conference on Data Mining, Berlin, Germany, Lecture Notes in Artificial Intelligence 6171, pp. 572-583, Berlin, Germany, July, 2010 (Scopus)

B Lecture Team

1. Summary

The lecture team includes the three members of the previous four ABI editions (MAP-I 2008/9, 2010/11, 2011/12 and 2012/13). **Manuel Filipe Santos (MFS)** and **Paulo Cortez (PC)** belong to the Intelligent Data Systems (IDS) - http://algoritmi.uminho.pt/research-teams/ids, research group of the Algoritmi R&D Centre (evaluated as "**Very Good**" by FCT), University of Minho. Both research on Adaptive Business Intelligence, Decision Support Systems and Data Mining. MFS performed his PhD in Distributed Learning Classifier Systems, while PC performed his PhD in Forecasting, Neural Networks and Evolutionary Optimization. **Rui Camacho (RC)** is from the Laboratory of Artificial Intelligence and Decision Support (LIAAD R&D centre, http://www.liaad.up.pt/**,** evaluated as "**Very Good**" by FCT), University of Porto. RC researches in Inductive Logic Programming and Data Mining. For more details, see the CVs in section B.3.

The team is willing to write didactic texts related to this unit. In effect, PC is currently discussing with Springer the writing of the book "Modern Optimization with R".

2. Coordinator

Manuel Filipe Santos (MFS)

3. CVs

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 associate professor at the Department of Information Systems, UMinho, teaching undergraduate and graduate classes of Business Intelligence and Decision Support Systems. He is also researcher at the Intelligent Data Systems (http://algoritmi.uminho.pt/research-teams/ids) 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] Tiago Miranda; António G Correia; Manuel F Santos; Luís R Sousa; Paulo Cortez NEW MODELS FOR STRENGTH AND DEFORMABILITY PARAMETERS CALCULATION IN ROCK MASSES USING DATA MINING TECHNIQUES ASCE's International Journal of Geomechanics, 2010 20-30.

[2] Ana Azevedo, Manuel Santos, A Perspective on Data Mining Integration with Business Intelligence in "Knowledge Discovery Practices and Emerging Applications of Data Mining: Trends and New Domains", IGI Global 2010, ISBN 978-1-60960-069-3.

[3] Henrique Santos; Manuel Filipe Santos; Wesley Mathew Supervised Learning Classifier System for Grid Data Mining Data Mining: Trends and New Domains, Viena, Intech, 2010, ISBN 978-953-7619-X-X.

[4] Júlio Duarte, Maria Salazar, Cesar Quintas, Manuel Santos, José Neves, António Abelha and José Machado Data Quality Evaluation of Electronic Health Records in the Hospital Admission Process, IEEE/ACIS 2010 Japan August, 2010 (ISI proceedings).

[5] Modelling intelligent behaviours in multi-agent based HL7 services, IEEE/ACIS 2010 Japan, August, 2010 (ISI proceedings).

[6] I. A. Iurgel; R. E. da Silva; M. F. dos Santos Towards virtual actors for acting out stories Edutainment 2010, China, August, 2010 (ISI proceedings).

[7] Manuel Filipe Santos; Wesley Mathew; Henrique Dinis Santos, Grid Data Mining by means of Learning Classifier Systems and Distributed Model Induction, GECCO 2011, July, Dublin.

[8] Filipe Portela, Manuel Filipe Santos, Marta Vilas-Boas, A Pervasive Approach to a Real-Time Intelligent Decision Support System in Intensive Medicine, book of Communications in Computer and Information Science, Springer-Verlag, 2011.

[9] Cabral, Alexandra; Abelha, António; Salazar, Maria; Quintas, César; Portela, Filipe; Machado, J.; Neves, José; Santos, M.F. Knowledge acquisition process for intelligent decision support in critical health care Information Systems and Technologies for Enhancing Health and Social Care, IGI Global Book -, 2013, (ISBN: 9781466636675)

[10] Santos, M.F.; Portela, Filipe; Miranda, Miguel; Machado, José, 1963-; Abelha, António; Silva, Álvaro; Rua, Fernando Grid data mining strategies for outcome prediction in distributed intensive care units Information Systems and Technologies for Enhancing Health and Social Care, IGI Global Book -, 2013, (ISBN: 9781466636675)

[11] Santos, M.F.; Azevedo, Ana Closing the Gap Between Data Mining and Business Users of Business Intelligence Systems: a Design Science Approach International Journal of Business Intelligence Research 3(4) - 2012.

[12] Neves, J., Ribeiro, J., Pereira, P., Alves, V., Machado, J., Abelha, A., Novais, P., Analide, C., Santos, M. Fernández-Delgado, M., Evolutionary intelligence in asphalt pavement modeling and quality-of information, in Journal of Progress in Artificial Intelligence, Vol 1, Springer-Verlag, 2012.

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

He participated in various R&D projects, being Principal Investigator of 3 projects:

- INTCARE II - Intelligent Decision Support System for Intensive Care, Principal Investigator, Approved for founding by FCT PTDC/EEI-SII/1302/2012, 2013-2014;
- GridClass – Learning Classifiers for Grid Data Mining, Principal Investigator, Approved for founding FCT GRID/GRI/81736/2006, 2008-2011;
- INTCARE - Intelligent Decision Support System for Intensive Care, Principal Investigator, Approved for founding FCT PTDC/EIA/72819/2006, 2008-2012.

Supervision of Graduate Students:

Supervised 15 MSc theses and 4 PhD theses. Currently he is supervising 9 PhD students.

Other relevant topics of his CV:

- **Co-organized** the EPIA 2007 – 13th Portuguese Conference on Artificial Intelligence.
- **Reviewer** of several conferences (e.g. AAMAS, EPIA, ICEIS, ICAART, MEDI, IARIA) and journals (e.g. European Journal of Operational Research, Intelligent Decision Making Support Systems);
- **Co-organizer of** the Ubiquitous Data Mining workshop of ECAI 2012, 2010, Knowledge Discovery and Business Intelligence - KDBI 2009, 2011 and 2013 and MASTA 2013 thematic tracks of EPIA; WISA/CISTI 2011 and Intelligent Systems/ESM 2011.

3.2 Paulo Cortez

Biography:

Paulo Cortez (PhD in Computer Science) is **Associate Professor** (with tenure) at the Department of Information Systems, University of Minho. He is also researcher at Algoritmi Centre, with interests in the fields of: Business Intelligence and Decision Support Systems; Data Mining and Machine Learning; Neural Networks and Evolutionary Computation; and Forecasting.

Relevant publications in the last 5 years:

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

[1] J. Peralta Donate, P. Cortez, G. Sánchez and A. de Miguel. Time series forecasting using a weighted cross-validation evolutionary artificial neural network ensemble. In **Neurocomputing**, Elsevier, In press (ISI impact factor 1.429).

[2] M. Stepnicka, **P. Cortez**, J.P. Donate, L. Stepnickova. Forecasting seasonal time series with computational intelligence: on recent methods and the potential of their combinations. In **Expert Systems with Applications**, Elsevier, 40(6):1981-1922, May 2013, ISSN 0957-4174 (ISI impact factor 2.203).

[3] P. Cortez and M.J. Embrechts. Using Sensitivity Analysis and Visualization Techniques to Open Black Box Data Mining Models. In **Information Sciences**, Elsevier, 225:1-17, March 2013, ISSN 0020-0255 (ISI impact factor 2.833).

[4] P. Cortez and J.P. Donate. Evolutionary Support Vector Machines for Time Series Forecasting. In A. Villa et al. (Eds.), Artificial Neural Networks and Machine Learning - ICANN 2012, **22nd International Conference on Artificial Neural Networks,** Lecture Notes in Computer Science 7553, pp. 523-530, Lausanne, Switzerland, September, 2012, Springer, ISSN 0302-9743, ISBN 978-3-642-33265-4 (ISI, Scopus)

[5] P. Cortez, M. Rio, M. Rocha and P. Sousa. Multiscale Internet Traffic Forecasting using Neural Networks and Time Series Methods. In **Expert Systems**, Wiley-Blackwell, 29(2):143-155, May 2012 (ISI impact factor 1.231).

[6] P. Cortez. Data Mining with Multilayer Perceptrons and Support Vector Machines. In D. Holmes and L. Jain (Eds.), **DATA MINING: Foundations and Intelligent Paradigms, Volume 2: Core Topics including Statistical, Time-Series and Bayesian Analysis**, ISRL 24, chapter 2, pp. 9-25, 2012. Springer (ISI).

[7] C. Lopes, P. Cortez, P. Sousa, M. Rocha and M. Rio. Symbiotic filtering for spam email detection. In **Expert Systems with Applications**, Elsevier, 38(8):9365-9372, August 2011 (ISI impact factor 1.924).

[8] M. Rocha, P. Sousa, P. Cortez and M. Rio. Quality of Service Constrained Routing Optimization using Evolutionary Computation, In **Applied Soft Computing**, Elsevier, Elsevier, 11(1):356-364, 2011 (ISI impact factor 2.415, Scopus)

[9] P. Cortez and M. Embrechts. Opening Black Box Data Mining Models Using Sensitivity Analysis. In Proceedings of the 2011 **IEEE Symposium on Computational Intelligence and Data Mining (CIDM)**, pp. 341-348, Paris, France, April, 2011 (ISI, Scopus).

[10] P. Cortez, Sensitivity Analysis for Time Lag Selection to Forecast Seasonal Time Series using Neural Networks and Support Vector Machines, In Proceedings of the **IEEE International Joint Conference on Neural Networks IJCNN**, pp. 3694-3701, Barcelona, Spain, July, 2010. (ISI, Scopus).

[11] P. Cortez, Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool, In P. Perner (Ed.), Advances in Data Mining, **Proceedings of 10th Industrial Conference on Data Mining**, Berlin, Germany, Lecture Notes in Artificial Intelligence 6171, pp. 572-583, Berlin, Germany, July, 2010 (ISI, Scopus)

[12] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In **Decision Support Systems**, Elsevier, 47(4):547-553, 2009. ISSN: 0167-9236. (ISI impact factor 1.873, Scopus, ACM Portal)

[13] R. Costa, N. Cachulo and P. Cortez. An Intelligent Alarm Management System for Large-Scale Telecommunication Companies. In L. Lopes et al. (Eds.), Progress in Artificial Intelligence, **14th Portuguese Conference on Artificial Intelligence (EPIA'2009)**, Lecture Notes in Artificial Intelligence 5816, pp. 386-399, Aveiro, Portugal, October, 2009. Springer. ISBN-10 3-642-04685-1. (34% acceptance rate, ISI, Scopus, DBLP)

[14] 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, 43 (3): 179–193, 2008 (ISI impact factor 1.882).

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

He participated in 4 R&D projects, being Principal Investigator of 1 project, 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.

Supervision of Graduate Students in the last 5 years:

Supervised 2 PhD thesis and 11 MSc thesis. Currently he is supervising 6 PhD students.

Other relevant topics of his CV:

- **Associate Editor** of the **Neural Processing Letters** journal (Springer, ISI), since 2008.
- **Guest Editor** of Knowledge Discovery and Business Intelligence issue of Expert Systems **journal** (ISI), in press.
- **Reviewer** of several ISI journals: Information Sciences, Data & Knowledge Engineering, Artificial Intelligence in Medicine, Computer Journal, Decision Support Systems, Neurocomputing, Expert Systems, Intelligent Data Analysis and Artificial Intelligence Communications.

- **Program Committee Member** of 66 int. conferences/workshops, such as: IDEAL 2013, IEEE CIDM 2013, ACM WIMS'13, ECAI2010, IEEE IJCNN 2012, IEEE FUZZ 2013, DDDM of IEEE ICDM 2011, PAKDD-09, AAMAS09, IEEE CBMS 2011.
- **Co-organizer** of 12 Workshops, such as: Machine Learning track of IBERAMIA 2012; Knowledge Discovery and Business Intelligence (KDBI EPIA 2013); Ubiquitous Data Mining (UDM-IJCAI 2013; UDM-ECAI 2012).
- **Invited lecturer** in the International Summer School of Neural Networks in Classification, Regression and Data Mining (2003-06;2008; 2010; 2012).
- Acted as **external examiner** of 14 MSc and 5 PhD thesis.
- **Author** of the open source RMiner library, which facilitates the use of Data Mining applications in R (http://www3.dsi.uminho.pt/pcortez/rminer.html).
- He is **vice-president** of the Portuguese Association for Artificial Intelligence (APPIA).
- He has a **strong post-graduate teaching experience**, having taught 24 MSc course units and 5 PhD course units in MAP-I Universities, University Institute of Lisbon, Polytechnic Institute of Bragança and Universitat Politècnica de Catalunya – BarcelonaTech.

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 Associate Professor at the Informatics Engineering Department of the Faculty of Engineering at UP, teaching undergraduate and graduate classes of Machine Learning and Data Mining. He is 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;
- Relational 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:

[1] Chapter 7 (pp 158-205) "Induction as a search procedure", Stasinos Konstantopoulos, Rui Camacho, Vítor Costa, Nuno Fonseca, in the book "Artificial Intelligence for Advanced Problem Solving Techniques", edited by Dimitris Vrakas e Ioannis Vlahavas, Aristotle University of Thessaloniki, Grece, pp. 166-216, 2008.

[2] Max Pereira, Nuno A. Fonseca Vítor Santos Costa Rui Camacho, "Interactive Discriminative Mining of Chemical Fragments", in Proceedings of the 19th International Conference on Inductive Logic Programming, Springer-Verlag, LNAI series, Florence, Italy, 2010.

[3] Nuno A. Fonseca, Fernando Silva, Rui Camacho and Ashwin Srinivasan, Parallel ILP for Distributed-Memory Architecture, in Machine Learning journal, Vol. 74, Number 3, pp. 257-279, March 2009

[4] N. Fonseca, V. S. Costa, R. Rocha, Rui Camacho, F. Silva,``Improving the Efficiency of ILP Systems", journal of Software: Practice and Experience, Vol. 39, Issue 2, pp. 189-219, Fev. 2009

[5] N. Fonseca, Camacho, R. Rocha, V. S. Costa, ``Compile the hypothesis space: do it once, use it often", Fundamenta Informaticae, Special Issue on Multi-Relational Data Mining(89):45-67, 2008.

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

He participated in various R&D projects (4 international projects), being Principal Investigator of 2 projects, namely:

ILP-Web-Service: An Inductive Logic Programming based Web service
and
Informatização de dados filogenéticos para avaliação da patogenicidade no DNA mitocondrial

Supervision of Graduate Students:

Has supervised 3 MSc theses, 10 *Mestrado Integrado* thesis and 5 PhD theses. Currently supervises 2 PhD students.

Other relevant topics of his CV:

- Co-organized the ILP 2004 – International Conference on Inductive Logic Programming.
- Co-organized the ECML/PKDD 2005 – European 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.
- Belongs to the editorial board of the International Journal of Computational Intelligence in Bioinformatics and Systems Biology (IJCIBSB)
- Made review work for the following international journals: Journal of Computational Intelligence (JCI); IEEE journal of Systems Man and Cybernetic (SMC-B); Data & Knowledge Engineering (DKE), journal of Artificial Intelligence Research (JAIR), Machine Learning journal, Journal of Information Science, The Computer Journal, Journal of Integrative Bioinformatics, Journal of Computational Intelligence and Bioinformatic (JCIB).
- Made review work for the book "Computational Intelligence in Bioinformatics", Springer-Verlag, 2007
- Is an "International Colaborator" of "N□ucleo de Apoio □a Pesquisa para Aprendizado de Máquina em Anáalise de Dados" of USP, Brasil.
- He was invited to evaluate 2 project proposals by the Research Council K.U.Leuven (Belgic) in 2011