

Editorial: Knowledge Discovery and Business Intelligence

Paulo Cortez¹ Manuel Filipe Santos¹

¹ Centro Algoritmi, Dep. Information Systems, University of Minho,
4800-058 Guimarães, Portugal

Email: pcortez@dsi.uminho.pt, mfs@dsi.uminho.pt

1 Introduction

Due to advances in Information Technology, nowadays it is easy to collect, store, process and share data. In effect, the amount of data stored by organizations or individuals is estimated to be growing exponentially [Lyman, 2003]. While Expert Systems (ES) were originally devised to solely mimic human experts [Buchanan, 1986], there is a pressure to extract as much useful information as possible from past data. Hence, the current trend to include data-driven models, possibly integrated with expert-driven models, into the decision-making process. Within this context, there are two relevant terms: Knowledge Discovery (KD) and Business Intelligence (BI). KD is a branch of the Artificial Intelligence (AI) field that aims to extract useful and understandable high-level knowledge from complex and/or large volumes of data [Fayyad et al., 1996]. BI is an umbrella term that represents several computer architectures, tools, technologies and methods (e.g., Data Warehousing, On-line analytical processing and KD) to access past data and support decision-making in public and corporate enterprises, from operational to strategic level [Turban et al., 2010].

KD and BI are faced with new challenges. For instance, recent communication technologies, such as WiMAX, lead to interesting network optimization problems, which can be solved using a KD approach. Also, when adopting fuzzy clustering methods, it is not clear how to correctly identify the ideal number of clusters. Moreover, a large effort has been put on static analysis of social networks. Yet, these social networks exist in time and thus may evolve under a dynamic environment. Furthermore, while there are several public dictionaries with synonymy information, the building of thesaurus often requires manual work and is usually incomplete.

This special issue, entitled “Knowledge Discovery and Business Intelligence” (KDBI), focuses on new KD approaches that aim to solve new challenges (such as previously described), leading to a potential valuable impact in several BI domains. This special issue consists of extended versions of papers from the 2nd KDBI workshop of the 15th Portuguese Conference on Artificial Intelligence (EPIA 2011) that was held in Lisbon, Portugal. A total of 27 papers were submitted to the 2nd KDBI workshop, from which the best 10 papers were invited for this special issue. Each extended paper was reviewed by a minimum of three reviewers (related with both

the 2nd KDBI workshop and Expert Systems journal), and passed through two rounds of reviews. Finally, the best four papers were accepted, corresponding to an acceptance rate of 15%, when considering the initial 2nd KDBI submitted papers, and 40%, when considering the extended invited papers. In the next section, we provide a brief introduction to these papers.

2 Contents of the special issue

For this special issue, four outstanding papers were selected. All these papers propose novel intelligent data analysis methods that are tested over real-world data and with an analysis of their valuable impacts in the business domain.

The paper of [Stolojescu et al., 2013] proposes an original KD methodology for the telecommunications domain. In particular, a Long Range Dependence (LRD) perspective is used to estimate the quality of base stations positioned within a wireless network that uses the new WiMAX communication technology. The proposed approach can be used to optimize the functioning of the WiMAX network (e.g., turning the traffic more uniform and allowing to accommodate an increased number of users). Moreover, it could also increase network security by detecting traffic anomalies.

The work of [Nascimento et al., 2013] presents further research of a novel fuzzy additive spectral clustering method, termed FADDIS. Such method has the advantage of providing guidance for choosing the number of clusters, while it compares favorably with state of the art methods. In particular, FADDIS is potentially useful when modeling pairwise similarities by additive properties (e.g., finding thematic clusters of the activities conducted in an organization).

The paper of [Oliveira and Gama, 2013] introduces a new approach for analyzing dynamic temporal changes in social networks, at both node-level and community level. In particular, three-order tensors are used to project the trajectories of social entities into a 2D space, which helps in understanding latent properties of the dynamic social network. While the approach was tested in friendship networks among university freshmen, there are also potential benefits in business applications, such as churn prediction and viral marketing.

The work of [Oliveira and Gomes, 2013] discusses a novel method for the enrichment of thesaurus by including information acquired automatically from dictionaries. While the method (which includes a clustering approach) was applied to enrich a Portuguese thesaurus with information from three dictionaries, it can be adapted to other languages. Such thesaurus is useful within several domains, including not only text mining and information retrieval (e.g., stemming) but also BI. For instance, it could help in the mapping of attributes with similar meaning but with different names, thus facilitating the extraction of data from multiple sources into a data warehouse.

All these papers represent the best contributions to the Knowledge Discovery and Business Intelligence workshop of EPIA 2011. Following the original purpose of the KDBI workshop, we hope that this special issue contributes for enrichment of the KD and BI fields, including in particular an increased interaction between KD and BI.

Acknowledgments

Many individuals contributed to this special issue. We would like to thank Dr. Jon G. Hall, Editor-in-Chief of Expert Systems, for his interest and support. We also wish to thank the other KDBI 2011 track (of EPIA) co-organizers, Nuno Marques, Luís Cavique and João Gama, without their help this issue would not have been possible. Finally, we thank the authors, who contributed with their papers, and the reviewers (from the KDBI 2011 program committee and also others), who contributed with their expert and valuable opinions.

References

- [Buchanan, 1986] Buchanan, B. G. (1986). Expert systems: working systems and the research literature. *Expert systems*, 3(1):32–50.
- [Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). *Advances in Knowledge Discovery and Data Mining*. MIT Press.
- [Lyman, 2003] Lyman, P. (2003). How much information? <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/>.
- [Nascimento et al., 2013] Nascimento, S., Felizardo, R., and Mirkin, B. (2013). Laplacian normalization for deriving thematic fuzzy clusters with an additive spectral approach. *Expert systems*, pages –.
- [Oliveira and Gomes, 2013] Oliveira, H. and Gomes, P. (2013). Towards the automatic enrichment of a thesaurus with information in dictionaries. *Expert systems*, pages –.
- [Oliveira and Gama, 2013] Oliveira, M. and Gama, J. (2013). Visualization of evolving social networks using actor-level and community-level trajectories. *Expert systems*, pages –.
- [Stolojescu et al., 2013] Stolojescu, C., Lenca, P., Moga, S., and Isar, A. (2013). Wimax traffic analysis and base stations classification in terms of lrd. *Expert systems*, pages –.
- [Turban et al., 2010] Turban, E., Sharda, R., Aronson, J., and King, D. (2010). *Business Intelligence, A Managerial Approach*. Prentice-Hall, 2nd edition.

The authors

Paulo Cortez

Paulo Cortez is Associate Professor at the Department of Information Systems, University of Minho, Portugal. He is also researcher at the Intelligent Data Systems group of Centro Algoritmi, with interests in the fields of: Business Intelligence (Decision Support, Data Mining and Forecasting); and Artificial Intelligence (Neural

Networks, Evolutionary Algorithms and Applications). Currently, he is associate editor of the *Neural Processing Letters* journal and participated in 9 R&D projects (principal investigator in 2). He is co-author of more than eighty indexed (ISI, Scopus) publications in international journals (e.g., Expert Systems, Information Sciences) and conferences. Web-page: <http://www3.dsi.uminho.pt/pcortez>

Manuel Filipe Santos

Manuel Filipe Santos is Associate Professor at the Department of Information Systems, University of Minho, Portugal, and leader of the Intelligent Data Systems group of Centro Algoritmi, with interests in the fields of: Business Intelligence, Data Mining and Learning Classifier Systems. He participated in several R&D projects (principal investigator in 3). He is also co-author of several publications in international journals (e.g., Artificial Intelligence in Medicine) and conferences (e.g., GECCO).