Preface

Alfredo Cuzzocrea

ICAR-CNR and University of Calabria,
Via P. Bucci, 41C, Rende,
87036 Cosenza, Italy
E-mail: cuzzocrea@si.deis.unical.it

Biographical notes: Alfredo Cuzzocrea is actually a researcher at the Institute of High Performance Computing and Networking of the Italian National Research Council, Italy, and an Adjunct Professor at the Department of Electronics, Computer Science and Systems of the University of Calabria, Italy. His research interests include multidimensional data modelling and querying, data stream modelling and querying, data warehousing and OLAP, OLAM, XML data management, web information systems modelling and engineering, knowledge representation and management models and techniques, grid and P2P computing. He is author or co-author of more than 95 papers in referred international conferences (including EDBT, SSDBM, ISMIS, ADBIS, DEXA, DaWaK, DOLAP, IDEAS, SEKE, WISE, FQAS, SAC) and international journals (including DKE, JIIS, IJDWM, WIAS). He serves as programme committee member of referred international conferences (including ICDM, SDM, PKDD, PAKDD, CIKM, ICDCS, ER, WISE, DASFAA, FQAS, SAC) and as review board member of referred international journals (including TODS, TKDE, TSMC, IS, DKE, JIIS, IPL, TPLP, COMPJ, DPDB, KAIS, INS, IJSEKE, FGCS). He also serves as PC Chair in several international conferences and as Guest Editor in international journals like JCSS, DKE, KAIS, IJBIDM, IJDMMM and JDIM.

This special issue on "OLAP Intelligence: Meaningfully Coupling OLAP and Data Mining Tools and Algorithms" of *International Journal of Business Intelligence and Data Mining* focuses on latest research results and open research challenges on the problem of effectively and efficiently coupling OnLine Analytical Processing (OLAP) and Data Mining tools and algorithms towards a new scientific discipline known under the term OnLine Analytical Mining (OLAM), which represents one of the emerging research topics of next years in the context of knowledge discovery methodologies.

Nowadays, it is widely recognised that OLAP technology provides powerful analysis tools for extracting useful knowledge from large amounts of data stored in different and highly heterogeneous formats, and very often distributed across networked settings ranging from conventional wired environments to innovative wireless and Peer-to-Peer networks. Several advantages confirm the benefits coming from the OLAP analysis model:

- the amenity of 'naturally' representing real-life data sets that are multilevel, multidimensional and highly correlated in nature
- the amenity of analysing multidimensional data according to a multi-resolution and multi-granularity vision
- the rich availability of a wide class of powerful OLAP operators and queries
- the integration of OLAP with more complex analysis tools coming from statistics, time series analysis and data mining.

OLAM, introduced by Jiawei Han in his seminal paper in 1997, is an elegant and successful solution to the research challenges mentioned earlier, and consists in meaningfully combining the powerful of OLAP with the effectiveness of Data Mining tools and algorithms capable of discovering interesting knowledge from large amounts of data (e.g., the data cell set of a given OLAP data cube) by means of clustering, classification, association rule discovery, frequent item set mining, and so forth.

During the last decade, researchers have devoted their attention on the issue of meaningfully coupling OLAP and Data Mining tools and algorithms, leading to the term *OLAP Intelligence*. This great interest is essentially due to both exciting theoretical perspectives, such as complexity issues of executing time-consuming Data Mining routines over very large OLAP data cubes and relevant application/practical issues, which have a great impact in a plethora of real-life scenarios ranging from conventional distributed database management systems and cooperative information systems to innovative data stream management systems and sensor network data analysis tools.

Indeed, OLAP-Intelligence/OLAM is a novel and, actually, less explored research area, and several research questions still remain open. Contrary to this, more mature scientific disciplines like Data Mining and Data Warehousing are well developed and enriched by leading research results in the vest of elegant, effective and highly efficient models, techniques and algorithms for knowledge representation, management and discovery from massive amounts of data. This also constitutes a source of invaluable (research-) knowledge for next-generation efforts in the field.

With the aim of adequately fulfilling both theoretical and practical gap of the so-relevant OLAP-Intelligence/OLAM research area, this special issue contains six papers, which have gone through two rigorous review rounds before being accepted for the final inclusion.

The first paper, titled 'Mining significant change patterns in multidimensional spaces', by Ronnie Alves, Joel Ribeiro and Orlando Belo, proposes a new OLAP Mining method for exploring interesting trend patterns, with the main goal of mining the most TOP-K *significant changes*, in Multidimensional Spaces (MDS). The proposed method is carried out via applying a *gradient-based cubing strategy*. The challenge is then finding *maximum gradient regions*, which maximise the task of detecting TOP-K *gradient cells*. Authors also introduce several heuristics to prune MDS efficiently, as the search sizes of these spaces can become a critical bottleneck for OLAP Mining applications. Summarising the OLAP Mining method proposed by authors allows us to:

- evaluate significant changes in MDS by means of pushing *gradient search* into the MDS partitioning process
- measure gradient regions *spreadness* for data cubing

- measure *periodicity awareness* of a change, assuring that it is a change pattern and not just an isolated event
- devise a *Rank-Gradient-based Cubing* to mine significant change patterns in MDS. Finally, a wide and comprehensive experimental evaluation on synthetic multidimensional databases confirms the benefits of the proposed OLAP Mining method.

The second paper, titled 'Mining convergent and divergent sequences in multidimensional data', by Marc Plantevit, Anne Laurent and Maguelonne Teisseire, focuses the attention on the problem of devising Data Mining tools over OLAP data cubes able to take all specificities of OLAP data into account, e.g., *multidimensionality, multilevel hierarchies*, time. Inspired by this main motivation, authors propose a novel OLAP Mining method for discovering *multidimensional sequential patterns* from *multidimensional sequence databases* among several levels of (OLAP) hierarchies. The aim of these patterns is to describe inner trends of multidimensional sequences. In particular, authors define two types of multidimensional sequences:

- *convergent sequences*, where elements (i.e., multidimensional items) become more and more precise
- *divergent sequences*, where elements become instead more and more general.

To effectively and efficiently mine both classes of multidimensional sequences, authors propose a pattern-growth-based algorithm, whose performance is demonstrated via a campaign of experiments on both synthetic and real multidimensional sequence databases.

The third paper, titled 'Reduced representations of Emerging Cubes for OLAP database mining', by Sébastien Nedjar, Alain Casali, Rosine Cicchetti and Lotfi Lakhal, proposes an interesting approach to reduce the size of *Emerging Cubes*, which are specialised data structures for OLAP Database Mining. Given two *categorical database relations*, the Emerging Cube allows us to compare two data cubes built on top of these relations via capturing trends that are irrelevant in the data cube of a relation while being significant in the data cube of the other relation. To trend classification purposes, Emerging Cubes are represented via classical *borders*, which are the boundaries of the solution (classification) space and can support classification tasks to know whether a trend is emerging (positive example) or not (negative example). Despite this, as authors state, borders do not make possible to retrieve measure values and therefore cannot be used to answer OLAP queries over multidimensional data cubes. To fulfil this need, authors propose the *Emerging Closed Cube*, which is the most reduced and lossless representation of an Emerging Cube. In particular, two Emerging Closed Cube variants are introduced:

- **L**-*Emerging Closed Cube*, which is based on the concept of *cube closure*
- *Emerging Quotient Cube*, which is based on a revisited version of the well-known *Quotient Cube*, proposed to support data cube summarisation, thus exposing the amenity of preserving the 'specialisation/generalisation' property that characterises Quotient Cubes.

The latter property allows us to navigate the multidimensional and hierarchical space of an Emerging Cube by means of popular roll-up/drill-down OLAP operations. Authors complete their analytical contribution through a wide and comprehensive experimental analysis on synthetic and real-life data sets meant to measure the size of the two new proposed Emerging Cube variants in comparison with classical border-based Emerging Cube representations. This analysis clearly shows the appreciable space reduction owing to both **L**-Emerging Closed Cubes and Emerging Quotient Cubes with respect to Emerging Cubes, while retaining important OLAP functionalities like decision-support query answering over multidimensional data cubes (thanks to **L**-Emerging Closed Cubes) and hierarchy-based interactive navigation of multidimensional data cubes (thanks to Emerging Quotient Cubes).

In the fourth paper, titled 'Fragmenting very large XML data warehouses via *K*-means clustering algorithm', by Alfredo Cuzzocrea, Jérôme Darmont and Hadj Mahboubi, authors put emphasis on XML data sources, which are more and more gaining popularity in the context of a wide family of Business Intelligence (BI) and OLAP applications, owing to the amenities of XML in representing and managing semi-structured and complex multidimensional data. As a consequence, many *XML data warehouse models* have been proposed during past years to handle heterogeneity and complexity of multidimensional data in a way traditional relational data warehouse approaches fail to achieve. However, *XML-native database systems* currently suffer from limited performance, both in terms of volumes of manageable data and query response time. Therefore, recent research efforts are focusing the attention on *fragmentation techniques*, which are able to overcome the limitations discussed earlier. Inspired by this research challenge, authors propose the use of *K*-means clustering algorithm for effectively and efficiently supporting the fragmentation of very large XML data warehouses, and, at the same time, completely controlling and determining the number of originated fragments via adequately setting the parameter *K*. In support of their proposed XML data warehouse fragmentation framework, authors provide a comprehensive experimental assessment where they compare the efficiency of their proposal against those of classical *derived horizontal fragmentation algorithms*, which have been studied in the context of relational data warehouses, adapted to XML data warehouses.

The fifth paper, titled 'Embedded indicators to facilitate the exploration of a data cube', by Véronique Cariou, Jérôme Cubillé, Christian Derquenne, Sabine Goutier, Françoise Guisnel and Henri Klajnmic, addresses the issue of effectively and efficiently exploring OLAP data cubes via operators such as drill-down, roll-up or slice, which are widely used by business analysts as a decision-support tool. They state that, while exploring the target data cube, business analysts are rapidly confronted with analysing a huge number of drill paths according to the available different dimensions. Usually, business analysts are only interested in a small part of these paths, which correspond to either high statistical associations between dimensions or atypical cell values. Moreover, identifying the most interesting cells is a matter for business analysts. As a consequence, authors argue that coupling OLAP technologies and mining methods may help business analysts towards the automation of this tedious task. With these ideas in mind, author proposes a *discovery-driven method* to facilitate the whole process of exploring the data cube by identifying the most relevant dimensions to expand. At each step of the exploration process, a *built-in rank* on dimensions is displayed to the end-user, which is still free to choose the right dimension to expand for his or her analysis. Built-in rank on dimensions is performed through *indicators* computed on the fly according to the user-defined data selection. Finally, authors also show through a case study how this methodology offers an effective support to OLAP-based decision-making activities, in the vest of a tool that is directly integrated within a commercial OLAP management system.

In the sixth paper, titled 'A novel visualisation and interaction technique for exploring multidimensional data', by Lucio Ieronutti and Maurizio Pighin, authors start from recognising actual limitations of OLAP tools for visualising multidimensional data. They state that in actual implementations the result of an OLAP query over multidimensional data is typically displayed on the basis of a matrix format, where table rows and columns correspond to data cube dimensions, whereas table cells correspond to data cube measures. Some tools are able to visualise such results via adopting general *visualisation techniques*, which are often characterised by limited data interaction and exploration capabilities. To fulfil this of actual OLAP interfaces, authors propose an innovative *visual and interact technique* for the analysis of multidimensional data. In particular, the proposed solution is based on *three-dimensional data cube representations* that can be explored using dynamic queries and that combines colour-coding, detail-on-demand cutting planes and viewpoint control techniques, to enhance the whole visualisation process. Authors complete their analytical contribution via demonstrating the effectiveness of the proposed OLAP visualisation tool through some practical examples showing how it can be used for extracting useful information from real-life multidimensional data.

The Editor would like to thank very much the Editor-In-Chief of *International Journal of Business Intelligence and Data Mining*, Professor David Taniar, for accepting his proposal of a special issue focused on OLAP-Intelligence/OLAM, and for assisting him whenever required. The Editor would also like to thank all the reviewers who have worked within a tight schedule and whose detailed and constructive feedbacks to authors have contributed to substantial improvement in the quality of final papers. The complete International Program Committee of this special issue is listed next. Last but not least, the Editor is grateful to the authors who have submitted papers to this special issue. The Editor truly appreciates their patience and understanding throughout the review process.

Pedro Furtado, University of Coimbra, Portugal
Matteo Golfarelli, University of Bologna, Italy
Carlos Hurtado, University of Chile, Chile
Jens Lechtenborger, University of Munster, Germany
Jason Li, Drexel University, PA, USA
Patrick Martin, Queen’s University, Ontario, Canada
Rokia Missaoui, University of Quebec, Quebec, Canada
Muhesh Mohania, IBM India Research Lab, India
Mirek Riedewald, Cornell University, NY, USA
Timos Sellis, National Technical University of Athens, Greece
Alkis Simitsis, Stanford University, CA, USA
Manolis Terrovitis, The University of Hong Kong, China
Igor Timko, Free University of Bozen-Bolzano, Italy
Juan Trujillo, University of Alicante, Spain
Wei Wang, University of New South Wales, Australia
Robert Wrembel, Poznan University of Technology, Poland