Editorial

Juan Ramon Rabuñal*, Julián Dorado and Alejandro Pazos Sierra

Information and Communications Technologies Department,
School of Computer Science,
University of A Coruña,
Campus de Elviña s/n, 15071, A Coruña, Spain
E-mail: juanra@udc.es
E-mail: julian@udc.es
E-mail: apazos@udc.es
*Corresponding author

Biographical notes: Juan Ramon Rabuñal obtained his MS in Computer Science in 1999, PhD in Computer Science in 2002, and PhD in Civil Engineering in 2008, all of them at the University of A Coruña, Spain. Currently, he shares his time between his lecturer position (Associate Professor) at the Faculty of Computer Science and the direction of the Center of Technological Innovations in Construction and Civil Engineering of the University of A Coruña. He has also headed several research projects. His main research interests are artificial intelligence: artificial neural networks, evolutionary computation, artificial vision and the application of artificial intelligence to civil engineering.

Julián Dorado is an Associate Professor in the Faculty of Computer Science in the University of A Coruña. He finished his graduate studies in computer science in 1994. In 1999, he obtained his PhD, with a special mention of European Doctor. In 2004, he has finished his graduate studies in biology. He has worked as a teacher at the university for more than eight years. He has published many books and papers on several journals and international conferences. He is currently working on bioinformatics, evolutionary computing, artificial neural networks, computer graphics and data mining.

Alejandro Pazos Sierra is a Professor at University of A Coruña. He received his MS in Medicine from University of Santiago de Compostela in 1987, PhD in Computer Science from University Politécnica de Madrid, Spain in 1989, and PhD in Medicine from University Complutense de Madrid, Spain in 1996. He founded and is currently the Director of Artificial Neural Networks and Adaptive Systems Laboratory in the Faculty of Computer Science at A Coruña University. His current research interests include artificial neural networks, medical image, evolutionary computation, adaptive systems, control medical systems and telemedicine. He is the author of more than 40 published papers and is a member of the IEEE and several honour societies, including ACM, International Neural Network Society, American Association for the Advancement of Science, Internet Society, etc.

In recent years, the biological sciences have experienced an exponential growth in the creation of data as well as in the collective body of knowledge in this domain. This growth reflects the significant technological development in instrumentation aimed at

data collection. These advances have made it possible to conduct experiments with high throughput. Moreover, these technological developments have also resulted in lower associated costs, such that research teams can access these technologies in increasing numbers.

At the same time, the ever-increasing amounts of data lead to an exponential growth in research centred on 'omics': those areas that include genomics, transcriptomics, proteomics, metabolomics, and other types of large data ensembles that have been published as part of sizeable public databases and which can improve the understanding of biological processes. Clearly, an improved understanding of these processes depends on the ability of scientists to organise, integrate, share, and interpret the information in an effective manner.

In this context, the objective of translational research is to transform and transfer scientific discoveries that emerge in the laboratory, clinical environment or population studies, into clinical applications such that the incidence, morbidity and/or mortality rates of disease are reduced. This type of research entails constant interaction: a deliberate exchange of knowledge and resources that currently requires complex information systems that are not only interconnected but also coherently integrated. Translational research covers all phases of research: from the first stages of clinical investigations through environmental and societal actions, and includes the process of transferring data, knowledge, and technologies to meet various goals (including diagnostic and therapeutic decision-making as well as organisational and control decisions). It has been demonstrated that the efficient transfer of information and knowledge between the original data sources in the laboratory at one end, and at the other end, the care of the patient, requires that this transfer be a bidirectional and continuous process. Thus, the investigators involved in basic research should provide clinicians with new tools to better support patient care and treatment. These tools should additionally serve to evaluate the quality and impact on providing health care of using the tools themselves, and clinicians should engage in exploring innovative ways to study and follow the course of disease and outcomes, thereby feeding back into new avenues for basic investigation.

The analysis of biomedical information has become vital in understanding pathologies as well as the underlying biological processes. Such analysis is also essential to uncover possible relationships that may exist between large collections of data, and those that might shed light on the etiopathology of complex diseases.

In this broad context, this issue presents an analysis of biomedical data in different areas, from basic gene analysis through the study of proteomics and an examination of biomedical literature, to the analysis of clinical data. With this in mind, the issue is organised as follows.

The first paper, 'Clustering genome data based on approximate matching', by Patil et al., presents an approach for utilising clustering techniques such as k-means, fuzzy C-means (FCM), and probabilistic C-means (PCM) methods, to identify the species to which a particular sequence belongs given the sequence of two or three nucleotides as a starting point. The innovation of this approach lies in that exact sequence matching approaches are not used, but instead an approximate matching is pursued. This offers a higher degree of flexibility and permits the identification of sequences with a great deal of variability, as is the case with nucleotide sequences.

In the second article, 'Time-based discovery in biomedical literature: mining temporal links', Loglisci takes up one of the most difficult tasks in data mining, which is the exploitation or mining of textual data. This work seeks to interconnect concepts in the

biomedical literature with the added dimension of doing so in the temporal domain. The approach is based on a discretisation of the literature along the temporal domain by defining time intervals from which association rules can reveal conceptual interrelationships.

In the third article, 'Improvements to the quantum evolutionary clustering', Ramdane and Kholladi propose a new method of data clustering. Given a dataset the proposed algorithm optimises a measure of cluster quality to create a partition of the dataset. The algorithm described in this article is applied to two problems of classifying in clinical databases; the first one is a real problem in dermatology: The classification of differential diagnosis of erythemato-squamous diseases (psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris). The second one is another real world problem in clinical data mining, the classification (benign or malignant) of the breast cancer.

In the fourth article, 'Spot matching in 2D electrophoresis experiments', Fernandez-Lozano et al. describe an approximation approach to analyse proteomic image information. To conduct a comparative analysis of data contained in two proteomic images, it is necessary to identify each protein in one image with its homologous counterpart in the second image. The investigators employ a technique from computer vision called vision block matching, which consists of comparing image blocks to perform the matching. Once the corresponding proteins are identified in both images, it becomes possible to analyse the expressions of the proteins.

Finally, in their article 'Analysis of mixed C-means clustering approach for brain tumour gene expression data', Selva Kumar and Hannah Inbarani describe a clustering method to group genes that have similar functions, with a focus on genes associated with brain tumours. The authors compare three clustering techniques, fuzzy C-means, hierarchical clustering and mixed C-means, using two measures to validate the clusters (the Xie-Beni and DB indexes).

As suggested by these articles, this special issue attempts to offer a global view of the diverse approaches that are being considered in the analysis of information in very different areas in biomedicine, genomics, proteomics, and clinical data mining. In this regard, the editors hope that a new and potentially useful perspective has been presented to researchers in terms of emerging approaches to biomedical information analysis.

Finally, we wish to thank the authors for their contributions to this special issue, the reviewers for improving the overall quality of the articles, and the editors of the *International Journal of Data Analysis Techniques and Strategies* for giving us the opportunity to organise this special issue.