

Guest Editorial

Florentino Fdez-Riverola¹ / Miguel Rocha²

Selected Extended Papers of the 11th International Conference on Practical Applications of Computational Biology and Bioinformatics (PACBB)

¹ ESEI: Escuela Superior de Ingeniería Informática, Edificio Politécnico, University of Vigo, Ourense, Spain

² Centre of Biological Engineering, School of Engineering, University of Minho, Braga, Portugal, E-mail: mrocha@di.uminho.pt

DOI: 10.1515/jib-2017-0071

This special issue includes extended versions of a number of papers selected from the International Conference on Practical Applications of Computational Biology and Bioinformatics (PACBB 2017) that was held in Porto (Portugal) in June 2017. This forum, already in its eleventh edition, aims to gather and promote the interaction of a community of researchers developing applied Bioinformatics or Chemoinformatics solutions for diverse problems in biological and biomedical research.

The work presented at the conference covers a wide range of computer science and artificial intelligence methods, applied to address relevant problems arising from genome sequencing, omics data analysis, biological network reconstruction and analysis, systems biology or data integration, just to name a few topics of interest.

The five selected papers for this issue are globally focused on the topics of biomedical text mining and metabolomics data analysis and mining. These include recent challenges, such as the exploitation of metadata or the application of multivariate statistics, machine learning and deep learning methods over literature and omics data.

In the first paper by Antunes and Matos, the authors propose the use of word embeddings, calculated from the MEDLINE or the UMLS databases, as features in word disambiguation machine learning and knowledge-based algorithms, showing better results than previous approaches. The supervised approaches showed the best results in a common dataset for evaluating biomedical concept disambiguation.


The second paper, by the same authors, presents a supervised machine learning approach for identifying and ranking documents containing information about protein-protein interactions. In this work, the authors make use of a deep learning model, a convolutional recurrent neural network using word embeddings, being the method evaluated in one of the BioCreative III tasks.

The paper by Ferreira et al. proposes an automated tool to calculate measures of metadata quality, and suggest that this may be a mean to encourage data owners to increase the metadata quality of their submissions, contributing to higher quality data, enabling data sharing, and the overall accountability of publications. This tool is applied to the metabolomics database Metabolights, showing interesting results and highlighting lack of effort in the annotation tasks in some cases.

The final two papers in this issue are related to the analysis and mining of metabolomics data, in both cases using the R scientific computing system, through a package proposed by the authors, named *specmine*. The paper by Cardoso et al. addresses the use of Nuclear Magnetic Resonance (NMR) data to study banana peels in Southern Brazil, comparing their biochemical composition in different collection seasons along the year. A data analysis pipeline is proposed, using both univariate and multivariate data analysis tools, to address this research question.

Finally, the paper by Afonso et al., addresses the study of carotenoid contents in several cassava cultivars from Brazil. The paper proposes a machine learning pipeline to predict carotenoid contents from distinct types of data: CIELAB data (color measurements), Ultraviolet-visible (UV) data and a low-level data fusion of both. Results indicate good ability of the models to predict carotenoid concentrations from data obtained through low-cost techniques.

Miguel Rocha is the corresponding author.

 ©2017 Florentino Fdez-Riverola and Miguel Rocha, published by De Gruyter.

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.