Editorial

Current trends of granular data mining for biomedical data analysis

Biomedical data are available in many different formats, including numeric, textual reports, signals or images, and they come available from a variety of sources. Biomedical data typically suffer from incompleteness, uncertainty and vagueness, posing several challenges to perform data analysis, such high dimensionality, class imbalance or low numbers of samples [1,2]. Granular Computing, the term coined by Prof. L. A. Zadeh, provides a powerful tool for multiple granularity and multiple-view data analysis, which is of vital importance for understanding data driven analysis at different levels of ab-straction (granularity) [3]. It is worth stressing that human's capabilities in effective information organization and efficient reasoning with complex and uncertain information is highly dependent on hierarchical Granular Computing [4,5]. We have been witnessing significant advances of Granular Computing in the scientific and engineering domains. Data mining based on Granular Computing in biomedical data analysis is an emerging field which crosses multiple research disciplines and in-dustry domains. As a meta-mathematical methodology, granular data mining provides a theoretical framework for biomed-ical data analytics. It helps to extract knowledge when we are provided with an insufficient data that may also contain a significant amount of unstructured, uncertain and imprecise data. Granular data mining technology has exhibited some strong capabilities and advantages in intelligent data analysis and uncertainty reasoning for biomedical data. However, de-termining how to integrate Granular Computing and data mining to combine their advantages remains an interesting and important research topic. Recent survey indicated that granular data mining research has been focused on exploring the advantages, and also the challenges, derived from collecting and mining vast amounts of available biomedical data sources. It has therefore become strongly and timely justified to develop theoretical models and practical algorithms for carrying out granular data mining for biomedical data analysis.

This special issue aims at presenting a snapshot of recent theoretical foundations, algorithmic developments and appli-cations in the developments of granular data mining for biomedical data analysis. The papers in this special issue have been selected from a large number of submissions following a rigorous peer-review process. Twenty high-quality papers were finally accepted for publication after extensive reviews and revisions. This Special Issue brings well-focused, high-quality publications in this area with the intent to report on significant results and promote the visibility of granular data mining. They show some promising advancements in the area. In what follows, we provide a brief overview of the twenty impressive papers accepted here, separated into three groups according to the subject of their contribution.

The first group of eleven papers reported new developments on fuzzy sets and rough sets based granular data mining methodologies for biomedical data analysis.

The paper entitled “Evolutionary optimized fuzzy reasoning with mined diagnostic patterns for classification of breast tumors in ultrasound” by Qinghua Huang, Baozhu Hu, and Fan Zhang proposed a novel computer-aided-diagnostic algorithm for breast ultrasound data with human-in-the-loop, where user-participation was involved to manually give the scores for selected the Breast Imaging Reporting and Data System (BI-RADS) features.

The paper entitled “TOPSIS method based on a fuzzy covering approximation space: an application to biological nano-materials selection” by Kai Zhang, Jianming Zhan, and Yiyu Yao presented two pairs of covering-based fuzzy rough set models combining fuzzy neighborhood operators with fuzzy rough set models, revealed a new method for determining objective weights using the first pair of covering-based fuzzy rough set models, and provided a solution to solve the problem of bone transplant replacement materials selection.

The paper entitled “Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification” by Lin Sun, Xiaoyu Zhang, Yuhua Qian, et al. presented a novel feature selection method based on neigh-borhood rough sets using neighborhood entropy-based uncertainty measures for cancer classification from gene expression data. Experiments completed on ten gene expression datasets showed that the proposed algorithm is indeed efficient and outperformed other related methods in terms of the number of relevant selected genes and the classification accuracy.

The paper entitled ‘‘Local logical disjunction double-quantitative rough sets’’ by Yanting Guo, Eric C.C. Tsang, Weihua Xu, and Degang Chen put forward a local logical disjunction double-quantitative rough sets (LLDDRS) based on the importance, completeness and complementary natures of the relative and absolute quantitative information in describing approximation space, which provided an effective tool for discovering knowledge and making decisions in large data sets.

The paper entitled “PbFG: Physique-based fuzzy granular modeling for non-invasive blood glucose monitoring” by Weijie Liu, Anpeng Huang, Ping Wang, and Chao-Hisen Chu explored the use of granular computing in NGM modeling, and pro-posed the physique-based fuzzy granular modeling (PbFG) framework, in which a fuzzy physique classifier was applied to classify user's physique and then the blood glucose level was estimated by fusing corresponding BGC estimator outputs.

The paper entitled “Dual incremental fuzzy schemes for frequent item-sets discovery in streaming numeric data” by Hui Zheng, Peng Li, Qing Liu, et al. proposed two incremental FID schemes for numeric data: static fuzzy set partition and dynamic fuzzy set partition, both of which did not need to re-scan any previous batches of data. The experimental results demonstrated the computational efficiency of proposed method.

The paper entitled “A physiological and behavioral feature authentication scheme for medical cloud based on fuzzy-rough core vector machine” by Liming Fang, Changchun Yin, Lu Zhou, et al. proposed a physiological and behavioral feature authentication scheme based on fuzzy-rough theory to limit the access right of medical devices. This scheme required the doctor's own gesture for the authorization to access the medical device. The fuzzy-rough core vector machine (FRCVM) algorithm was adopted in this scheme to achieve high classification accuracy and efficiency. The results indicated that the solutions were highly secure and practical.

The paper entitled “A robust swarm intelligence-based feature selection model for neuro-fuzzy recognition of mild cog-nitive impairment from resting-state fMRI” by Ahmed M. Anter, Yichen Wei, Jiahui Su, et al. presented a new resting-state functional magnetic resonance imaging (rs-fMRI) data analysis approach (CBGWO-ANFIS) based on the chaotic binary grey wolf optimization and adaptive neuro-fuzzy inference system to identify the mild cognitive impairment patients (MCI) with Alzheimer's disease. The results indicated that the proposed CBGWO-ANFIS approach with the Chebyshev chaos map showed a higher accuracy, higher convergence speed, and shorter execution time than other chaos maps, and was a potential tool for early diagnosis of MCI.

The paper entitled “PARA: A Positive-region based Attribute Reduction Accelerator” by Peng Ni, Suyun Zhao, Xizhao Wang, et al. proposed an accelerator, based on fuzzy rough sets, for attribute reduction. The proposed accelerator was based on a strict mathematical foundation, since the monotonicity of key instance set and order preservation property was verified by mathematical reasoning. All these made sure that the reductions found by the accelerator were consistent with the non-accelerated algorithm.

The paper entitled “Segmentation of bias field induced brain MR images using rough sets and stomped-*t* distribution” by Abhirup Banerjee and Pradipta Maji, presented a novel method for simultaneous segmentation and bias field correction of brain MR images, which integrated the concept of rough sets and the merit of a recently introduced probability distribu-tion, called stomped-*t* (St-*t*) distribution. The proposed method employed both the expectation-maximization algorithm and hidden Markov random field model for accurate and robust image segmentation.

The paper entitled “Aggregation on ordinal scales with the Sugeno integral for biomedical applications” by Gleb Beli-akov, Marek Gagolewski and Simon James tackled the problem of learning fuzzy measure values in an ordinal framework for clinical applications. They formulated ordinal regression models for learning the Sugeno integrals from linguistic data and showed that the resulting piecewise linear objective functions can be decomposed into the difference of two convex functions, making them suitable for Difference of Convex (DC) optimisation approaches.

The second group of papers consists of three papers that focused on clustering methodologies based granular data mining for biomedical data analysis.

The paper entitled “Multiscale co-clustering for tensor data based on canonical polyadic decomposition and slice-wise factorization” by Zhenghong Wei, Hongya Zhao, Lan Zhao, and Hong Yan presented a theoretical framework to perform co-clustering for multidimensional data based on tensor and matrix decomposition. The extensive experimental data analysis on co-clustering spatial-temporal-genetic expression tensor in developing mouse brain and on co-clustering gene expression tensor of sclerosis patients under an IFN-β therapy demonstrated the validity and efficiency of the proposed co-clustering algorithm.

The paper entitled “A Bayesian possibilistic C-Means clustering approach for cervical cancer screening” by Fangqi Li, Shilin Wang, and Gongshen Liu proposed a Bayesian possibilistic C-means(BPCM) clustering algorithm for automatic cervical cancer screening that analyzed the related risk factors to provide preliminary diagnostic information for doctors.

The paper entitled “Multi-view cluster analysis with incomplete data to understand treatment effects” by Guoqing Chao, Jiangwen Sun, Jin Lu, et al. proposed an enhanced formulation for a family of multi-view co-clustering methods to cope with the missing data problem by introducing an indicator matrix whose elements indicated which data entries were ob-served and assessing cluster validity only on the observed entries. This approach was applied successfully to incomplete data collected in a heroin pharmacotherapy study.

The third group of six papers ventured into various granular data mining technologies for biomedical data analysis, including interval set, deep learning, and multi-part tensor decomposition.

The paper entitled “Spatio-temporal deep learning method for ADHD fMRI classification” by Zhenyu Mao, Yi Su, Guangquan Xu, et al. proposed a deep learning method called 4-D CNN based on granular computing that was trained based on the derivative changes in entropy. It calculated granularity at a coarse level by stacking layers, which sampled a subject’s rs-fMRI frames into several relatively short-term pieces with a fixed stride.

The paper entitled “Diabetic complication prediction using a similarity-enhanced latent dirichlet allocation model” by Shuai Ding, Zhenmin Li, Xiao Liu, et al. proposed a novel approach for diabetic complication prediction approach based on a similarity-enhanced latent Dirichlet allocation (seLDA) model, which was performed using the obtained similarity estima-tions as constraints. The experimental results showed that the proposed approach consistently outperformed the conven-tional approaches in both similarity estimations and diabetic complication predictions.

The paper entitled “A new hybrid wrapper TLBO and SA with SVM approach for gene expression data” by Alok Kumar Shukla, Pradeep Singh, and Manu Vardhan proposed a new hybrid wrapper approach to determine the optimal gene subsets from gene expression profiling. The approach, termed TLBOSA, used a combination of Teaching Learning-based Optimization (TLBO) and Simulated Annealing (SA) algorithm, which helped reveal the hidden patterns of tumors and enhance the inter-pretability of the selected genes. Experimental results and statistical analysis demonstrated that the proposed method could select discriminating input genes and achieve high classification accuracy.

The paper entitled “Deep feature learning for histopathological image classification of canine mammary tumours and hu-man breast cancer” by Abhinav Kumar, Sanjay Kumar Singh, K. Lakshmanan, et al. introduced a dataset of canine mammary tumour histopathological images and presented a preliminary study on automated binary classification of Canine Mammary Tumours (CMTs). The proposed framework resulted in a higher mean accuracy for binary classification of human breast cancer and CMTs, respectively.

The paper entitled “Uncertainty measures for interval set information tables based on interval δ-similarity relation” by Yimeng Zhang, Xiuyi Jia, Zhenmin Tang, et al. proposed several uncertainty measures for evaluating the granularity and uncertainty of interval set information tables, and a two-stage transferring algorithm to transform an incomplete single-valued information table to an interval set information table. The effectiveness of the proposed measures was validated on several biomedical datasets.

The paper entitled “Predicting tissue-specific protein functions using multi-part tensor decomposition” by Sameh K. Mo-hamed proposed a new method for predicting tissue-specific protein functions using tensor factorization on multi-part em-beddings, and applied tensor factorization to learn the scores for all possible protein-function associations for each studied tissue. Experimental evaluation indicated that the proposed model outperformed the state-of-the-art models in terms of the area under ROC and the precision-recall curve, respectively.

Finally, we would like to take this opportunity to express our sincere thanks to the authors who have shared their recent research results and to the anonymous reviewers whose critical, constructive comments have greatly improved the quality of the papers in this special issue. We wish to thank Information Sciences for giving us the opportunity to publish this special issue, and Elsevier for their efficient publication management. Special thanks should be given to the Editor-In-Chief of the journal, Witold Pedrycz, who worked closely with the Guest Editors to ensure this issue’s excellence and guarantee its success. This work has been carried out by the National Natural Science Foundation of China under Grant 61976120, the Natural Science Foundation of Jiangsu Province under Grant BK20191145, the Six Talent Peaks Project of Jiangsu Province under Grant XYDXXJS-048, and sponsored by Qing Lan Project of Jiangsu Province.

Weiping Ding*
Nantong University, China

Chin-Teng Lin
University of Technology Sydney, Australia

Alan Wee-Chung Liew
Griffith University, Australia

Isaac Triguero
University of Nottingham, United Kingdom

Wenjian Luo
University of Science and Technology of China

*Corresponding author.
E-mail address: *dwp9988@hotmail.com* (W. Ding)

References

[1] C.-H. Fung, Z.T. Ho Tse, K.-W. Fu, Converting big data into public health, Science, 347, 2015, p. 620. -620.

[2] N. Mehta, A. Pandit, Concurrence of big data analytics and healthcare: A systematic review, International Journal of Medical Informatics 114 (2018) 57–65.

[3] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 19 (1997) 111–127.

[4] J. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: Perspectives and challenges, IEEE Transactions on Cybernetics 43 (6) (2013) 1977–1989.

[5] W. Pedrycz, Granular computing for data analytics: a manifesto of human-centric computing, IEEE/CAA Journal of Automatica Sinica 5 (6) (2018) 1025–1034.