

**EDITORIAL****IEEE ACCESS SPECIAL SECTION EDITORIAL:  
DATA MINING AND GRANULAR COMPUTING  
IN BIG DATA AND KNOWLEDGE PROCESSING**

Data mining has actively contributed to solving many real-world problems with a variety of techniques. Traditional approaches in this field are classification, clustering and regression. During the last few years a number of challenges have emerged, such as imbalanced data, multi-label and multi-instance problems, low quality and/or noisy data or semi-supervised learning, among others [item 1) in the Appendix]. When these non-standard scenarios are encountered in the realm of big data, it remains an uncharted research territory, although a growing effort has been made to break the limits. The current trend is to address the classical and newly emerging data mining problems in big data and knowledge processing. Granular computing provides a powerful tool for multiple granularity and multiple-view data analysis at different granularity levels, which has demonstrated strong capabilities and advantages in intelligent data analysis, pattern recognition, machine learning and uncertain reasoning [item 2) in the Appendix]. Big data often contains a significant amount of unstructured, uncertain and imprecise data. There are new challenges regarding the scalability of granular computing when addressing very big data sets [item 3) in the Appendix]. Big data mining relies on distributed computational strategies; it is often impossible to store and process data on one single computing node. The exploration of data mining and granular computing in big data and knowledge processing is an emerging field which crosses multiple research disciplines and industry domains, including transportation, communications, social network, medical health, and so on.

The goal of this Special Section in IEEE ACCESS is to provide an opportunity to review the state-of-the-art of the recent data mining and granular computing in big data and knowledge processing, and bring together researchers in the relevant areas to discuss the latest progress, new research methodologies and potential research topics. We received a total of 45 submissions, and after a rigorous review process, 15 high-quality articles have been selected for publication, which are briefly discussed next.

Dimension reduction plays an important role in practical big data analysis and knowledge processing applications. To deal with the high-dimensional D-optimal design challenge, Xu, *et al.*, in the invited article entitled “Finding high-dimensional D-optimal designs for logistic models via

differential evolution,” proposed the NovDE algorithm to solve high-dimensional locally D-optimal design of logistic models with and without factor interactions, where a new novelty-based mutation strategy is adopted to explore various regions of the search space so that the diversity of the population would be preserved. The new novelty-based mutation strategy is collaborated to balance exploration and exploitation well at early or medium stages of evolution so that both convergence and diversity of the population are enhanced, and the premature convergence issue is alleviated. This data mining technology in evolutionary computation can find a highly efficient D-optimal design for high-dimensional real-world applications. To extract significant features from high dimensional data and to rapidly provide satisfactory clustering results, Sun, *et al.*, in the article entitled “An affinity propagation clustering method using hybrid kernel function with LLE,” proposed an efficient LLE-HKAP algorithm for high dimensional gene expression datasets and standard UCI datasets, where the first part reduces the dimensions of high-dimensional data and retains only the most significant data with the LLE algorithm, and the second part investigates a novel HKAP algorithm with a new global kernel. Zhang *et al.*, in the article entitled “CPCA: A feature semantics based crowd dimension reduction framework,” proposed a crowd-based dimension reduction framework called Crowd Principal Component Analysis (CPCA). The aim is to reduce the computational overhead of traditional principle component analysis leveraging collective intelligence so as to bring dimension reduction to individuals who have limited computational resources. CPCA asks crowd workers to estimate the independence among the features based on their semantics.

The classification techniques are of vital importance for understanding current intelligent data analysis. To resolve some challenges of architectural style classification, Zhao, *et al.*, in the article entitled “Architectural style classification based on feature extraction module,” proposed a feature extraction module based on DNN and IEP method. DNN can learn high-layer features from the architectural style images. IEP was based on the local-consistency and the exotic-inconsistency assumptions that can find common features of the same style and differences among 25-class architectural styles. The new features were used to classify

the architectural style. To improve the prediction accuracy of real-time vehicle classification in intelligent transport systems, Chen, *et al.*, in the article entitled “A novel model based on AdaBoost and deep CNN for vehicle classification,” proposed a novel model based on AdaBoost algorithm and deep convolutional neural networks for vehicle classification, where a high-efficiency deep CNN model is designed to directly extract the features of vehicle images and the output layer of the CNN is taken as the base learner of the AdaBoost algorithm. To efficiently handle the large-scale and highly complex Chinese judgment documents, Ma, *et al.*, in the article entitled “An ontology driven knowledge block summarization approach for chinese judgment document classification,” proposed an ontology driven knowledge block summarization approach to computing document similarity for Chinese judgment document classification, where the lightweight domain ontologies are built by reusing and adapting the existing public categories of Chinese judgment documents and the WMD-based similarity computation was made for KNN based document classification. To address data streams with concept drift and class imbalance, Zhang, *et al.*, in the article entitled “Online active learning paired ensemble for concept drift and class imbalance,” proposed a novel online active learning paired ensemble for data streams with concept drift and class imbalance, which have two classifiers: one was a stable classifier to learn the entire data stream; the other was a dynamic classifier to be updated and replaced during the entire learning process to capture the sudden change in the data stream. To deal with a large number of medial datasets collected from IoT-based platform, Yang *et al.*, in the article entitled “GAN-based semi-supervised learning approach for clinical decision support in health-IoT platform,” proposed a novel semi-supervised learning approach in association with Generative Adversarial Networks (GAN) for supporting clinical decision making in IoT-based health service system, where GAN is adopted to not only increase the number of labelled data, but also compensate the imbalanced labelled classes with additional artificial data in order to improve the semi-supervised learning performance.

Based on granular computing, in particular on fuzzy sets techniques, Beliakov, *et al.*, in the article entitled “Fuzzy connectives for efficient image reduction and speeding up image analysis,” proposed an image reduction algorithm capable of differentiating between fine image details and noise in the image, particularly salt and pepper noise. The mode-like averaging functions as local reduction operators are better at preserving the essential characteristics of an image and can differentiate between salt and pepper noise and 1-pixel wide lines and curves. Bharill, *et al.*, in the article entitled “A generalized enhanced quantum fuzzy approach for efficient data clustering,” proposed an enhanced quantum-inspired evolutionary fuzzy C-means (EQIE-FCM) algorithm for clustering of data, which can perform the clustering of datasets by evolving some important parameters in several generations using the quantum computing concept. These parameters

are evolved in each generation using five major operations: Representation of the fuzzifier parameter and the cluster centers in the quantum bit, transformation process, formulation of the local and the global fitness function and quantum update function. The proposed algorithm outperformed the other evolutionary clustering algorithms and acquired promising results.

Moreover, this Special Section also covers recent advances of data mining applications in many other important topics. To resolve information filtering of massive digitization of network data, Chen, *et al.*, in the article entitled “Extracting the information backbone based on personalized time window,” proposed a new information backbone extraction strategy based on a time window and constructed a new adjacency matrix from the information backbones of users instead of the complete information. To overcome the parameter-sensitivity, the rough-outlier, and the unidentified-hub problems in community detection, Meng, *et al.*, in the article entitled “A modified distance dynamics model for improvement of community detection,” proposed a robust distance dynamics model, namely, Attractor++, which used a dynamic membership degree to determine the influence of exclusive neighbors on the distance instead of setting the cohesion threshold. To improve the productivity and reduce cost of the welding robot, Tong, *et al.*, in the article entitled “Research on intelligent welding robot path optimization based on GA and PSO algorithms,” applied the GA and discrete PSO algorithm to the welding robot path optimization, where the operator selection of GA is improved as well as the PSO algorithm with improved inertia weight determination. Meanwhile, both “swap operator” and “swap sequence” are introduced into the solution of the discrete problem of welding robot path planning optimization. To assign workers to tasks according to their skills on road networks in a real-time scenario, Liang, *et al.*, in the article entitled “Online multi-skilled task assignment on road networks,” proposed a practical problem, called Online Multi-Skilled Task Assignment on Road Networks (OMTARN) problem, and designed a framework of using batch-based algorithms (fixed batch-based algorithm and dynamic batch-based algorithm) to achieve an approximate solution. Deng, *et al.*, in the article entitled “A multimodel fusion engine for filtering webpages,” proposed the MMFEFWP architecture for the filtering of large datasets of webpages crawled from websites, and the multimodel fusion engine for webpage filtering can extract target webpages using multiple models, including models based on strings, trees, link ratios, similarity degrees, and tags and attributes.

We recognize that the above articles cannot cover all the aspects of data mining and granular computing in big data and knowledge processing. We hope that through the selection of the articles, further in-depth discussions and new contributions to the related areas can be stimulated.

Finally, we would like to thank all the authors for contributing their work to this Special Section, and all the reviewers for their timely and professional reviews, without which this

Special Section would not have been published as scheduled. We would also like to thank the Managing Editor, Dr. Bora M. Onat, the Publications Editors, and the Editor-in-Chief and Staff Members for their supportive guidance during the entire process in the organization of this Special Section.

## APPENDIX

### RELATED WORKS

- 1) X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97–107, Jan. 2014.
- 2) J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: Perspectives and challenges," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1977–1989, Dec. 2013.
- 3) H. V. Jagadish *et al.*, "Big data and its technical challenges," *Commun. ACM*, vol. 57, no. 7, pp. 86–94, Jul. 2014.

### WEIPING DING

Nantong University  
Nantong 226019, China  
e-mail: [dwp9988@hotmail.com](mailto:dwp9988@hotmail.com)

### GARY G. YEN

Oklahoma State University  
Stillwater, OK 74078-5032, USA  
e-mail: [gyen@okstate.edu](mailto:gyen@okstate.edu)

### GLEB BELIAKOV

Deakin University  
Burwood, VIC 3125, Australia  
e-mail: [gleb@deakin.edu.au](mailto:gleb@deakin.edu.au)

### ISAAC TRIGUERO

University of Nottingham  
Nottingham NG7 2RD, U.K.  
e-mail: [isaac.triguero@nottingham.ac.uk](mailto:isaac.triguero@nottingham.ac.uk)

### MAHARDHIKA PRATAMA

Nanyang Technological University  
Singapore 639798  
e-mail: [mpratama@ntu.edu.sg](mailto:mpratama@ntu.edu.sg)

### XIANGLIANG ZHANG

King Abdullah University of Science and Technology  
Thuwal 23955, Saudi Arabia  
e-mail: [xiangliang.zhang@kaust.edu.sa](mailto:xiangliang.zhang@kaust.edu.sa)

### HONGJUN LI

Nantong University  
Nantong 226019, China  
e-mail: [lihongjun@ntu.edu.cn](mailto:lihongjun@ntu.edu.cn)



**WEIPING DING** (M'16–SM'19) received the Ph.D. degree in computation application from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2013. He was a Visiting Scholar with the University of Lethbridge (UL), AB, Canada, in 2011. From 2014 to 2015, he was a Postdoctoral Researcher with the Brain Research Center, National Chiao Tung University (NCTU), Hsinchu, Taiwan. In 2016, he was a Visiting Scholar with the National University of Singapore (NUS), Singapore. From 2017 to 2018, he was a Visiting Professor with the University of Technology Sydney (UTS), Ultimo, NSW, Australia. He is currently an Associate Professor with Nantong University, China. He has published over 50 papers in flagship journals and conference proceedings as the first author. To date, he has held 14 approved invention patents. His main research directions involve granular computing, data mining, evolutionary computing, machine learning, and big data analytics. He is a member of the IEEE-CIS, ACM, and IAENG, and a Senior Member of CCF. He is a member of the IEEE SMC Granular Computing (GrC) Technical Committee, the IEEE SMC Soft Computing Technical

Committee, and the IEEE CIS Task Force on Adaptive and Evolving Fuzzy Systems. He was a recipient of the Computer Education Excellent Paper Award (First-Prize) from the National Computer Education Committee of China, in 2009. He was an Excellent-Young Teacher (Qing Lan Project) of Jiangsu Province, in 2014, and a High-Level Talent (Six Talent Peak) of Jiangsu Province, in 2016. He was awarded the Best Paper of ICDMA'15. He was a recipient of the Medical Science and Technology Award (Second Prize) of Jiangsu Province, China, in 2017, and the Education Teaching and Research Achievement Award (Third Prize) of Jiangsu Province, China, in 2018. He was awarded two Chinese Government Scholarships for Overseas Studies, in 2011 and 2016. He currently serves on the Editorial Advisory Board of *Knowledge-Based Systems*. He serves/served as an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Information Sciences*, *Swarm and Evolutionary Computation*, and IEEE ACCESS, and a leading Guest Editor for four international journals.



**GARY G. YEN** (S'87–M'88–SM'97–F'09) received the Ph.D. degree in electrical and computer engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1992. He was with the Structure Control Division, U.S. Air Force Research Laboratory, Albuquerque, NM, USA. In 1997, he joined Oklahoma State University, Stillwater, OK, USA, where he is currently a Regents Professor with the School of Electrical and Computer Engineering. His current research interests include intelligent control, computational intelligence, conditional health monitoring, signal processing, and their industrial/defense applications. He was a recipient of the Andrew P. Sage Best Transactions Paper Award from the IEEE Systems, Man, and Cybernetics Society, in 2011, and the Meritorious Service Award from the IEEE Computational Intelligence Society, in 2014. He served as the General Chair for the 2003 IEEE International Symposium on Intelligent Control held in Houston, TX, USA, and the 2006 IEEE World Congress on Computational Intelligence held in Vancouver, BC, Canada. He was an Associate Editor of the *IEEE Control Systems Magazine*, the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY,

*Automatica*, *Mechantronics*, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—Part B: CYBERNETICS, and the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, and the IEEE TRANSACTIONS ON CYBERNETICS. He was the Founding Editor-in-Chief of the *IEEE Computational Intelligence Magazine*, from 2006 to 2009. He served as the Vice President for the Technical Activities, from 2005 to 2006, and then the President of the IEEE Computational Intelligence Society, from 2010 to 2011.



**GLEB BELIAKOV** (SM'08) received the Ph.D. degree in physics and mathematics in Moscow, Russia, in 1992. He is currently a Professor with the School of Information Technology, Deakin University. His research interests are in the areas of fuzzy systems, aggregation operators, multivariate approximation, global optimization, and computational mathematics. He has authored over 150 research papers and two monographs in the mentioned areas, and a number of software packages. He served as an Associate Editor for the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Fuzzy Sets and Systems*, and TOP journals.



**ISAAC TRIGUERO** received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 2009 and 2014, respectively. He has been an Assistant Professor of data science, since 2016. He has published more than 60 international publications in the fields of big data, machine learning, and optimization (H-index=21 and more than 1600 citations on Google Scholar). His work is mostly concerned with the research of novel methodologies for big data analytics. He is a Section Editor-in-Chief of *Machine Learning and Knowledge Extraction* and an Associate Editor of *Big Data and Cognitive Computing* and IEEE ACCESS. He has acted as the Program Co-Chair of the IEEE Conference on Smart Data (2016), the IEEE Conference on Big Data Science and Engineering (2017), and the IEEE International Congress on Big Data (2018). He is currently leading a KTP project funded by Innovative UK and the energy provider E.ON that investigates Smart Metering data.



**MAHARDHIKA PRATAMA** received the Ph.D. degree from the University of New South Wales, Australia, in 2014. He was with the Centre of Quantum Computation and Intelligent System, University of Technology, Sydney, as a Postdoctoral Research Fellow of the Australian Research Council Discovery Project. He was a Lecturer with the Department of Computer Science and IT, La Trobe University, from 2015 to 2017. He is currently a Tenure-Track Assistant Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests involve autonomous deep learning, data stream mining, lifelong learning, decision and control of autonomous vehicle, and prognostic health management. He received several awards in the past five years, namely the 2018 IEEE TFS Outstanding Publication Award and the 2019 Amity University Researcher Award. He currently serves as the Editor-in-Chief of the *International Journal of Business Intelligence and Data Mining*.



**XIANGLIANG ZHANG** received the M.S. and B.S. degrees from Xi'an Jiaotong University, China, in 2006 and 2003, respectively, and the Ph.D. degree in computer science from INRIA-Universite Paris-Sud, France, in 2010. She was a European ERCIM Research Fellow with the Norwegian University of Science and Technology (NTNU), Norway, in 2010. She is currently an Associate Professor of computer science and directs the Machine Intelligence and Knowledge Engineering Group, KAUST, Saudi Arabia. She has published over 100 research papers in referred international journals and conference proceedings, including TKDE, SIGKDD, VLDB J, AAI, IJCAI, ICDM, ECML/PKDD, CIKM, and InfoCom. Her research mainly focuses on learning from complex and large-scale streaming data. She is the Reviewer of TKDE, TKDD, VLDB J, *Information Science*, DMKD, and KAIS. She regularly serves on the Program Committee for premier conferences, such as SIGKDD, ICDM, AAI, IJCAI, NIPS, and ICML. She was selected and invited to deliver an Early Career Spotlight talk at IJCAI-ECAI 2018.



**HONGJUN LI** received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2011. He was a Visiting Scientist with the Centre for Pattern Recognition and Machine Intelligence (CENPARMI), Concordia University, Montreal, QC, Canada. He has published over 40 papers in flagship journals and conference proceedings as the first author. His research interests include image processing and pattern recognition. He is a Senior Member of CCA, CCF, and CSIG. He is currently an Associate Professor with Nantong University, China.

...