Guest Editorial Special Issue on Recent Advances in Theory, Methodology, and Applications of Imbalanced Learning

IMBALANCED learning is a challenging task in machine learning, faced by practitioners, and intensively investigated by researchers from a wide range of communities. However, as pointed out in the book titled "*Imbalanced Learning: Foundations, Algorithms, and Applications*" and collectively authored by experts in the field, many if not most of the approaches to imbalanced learning are heuristic and *ad hoc* in nature, hence leaving many questions unanswered. To fill this gap, the aim of this Special Issue is to collect recent research works that focus on the theory, methodology, and applications of imbalanced learning. After carefully reviewing a large number of submissions, we selected 15 works to be included in this Special Issue. These works can be roughly categorized into three types: deep-learning-based methods (6), methods based on other machine-learning paradigms (7), and empirical comparative studies (2).

The first group of articles includes the following six studies that propose new deep-learning-based methods for imbalanced learning:

1) *Learning Deep Landmarks for Imbalanced Classification:* Bao *et al*. proposed a deep imbalanced learning framework called DELTA, which introduces the new concept of rebalancing samples in a deeply transformed latent space. DELTA conducts feature learning, sample rebalancing, and discriminative learning in a joint and end-to-end manner, providing the possibility to conduct imbalanced learning with structured feature extraction.
2) *Learning Multi-Level Density Maps for Crowd Counting:* Jiang *et al*. developed a multilevel convolutional-neural network (MLCNN) model to accurately estimate the crowd count from a given scene with imbalanced people distribution. They also introduce a new loss function called balanced loss and a new dataset of more than 1k images with about 50k head annotations.
3) *Objective Video Quality Assessment Combining Transfer Learning with CNN:* Zhang *et al*. proposed a full-reference (FR) video quality assessment (VQA) metric integrating transfer learning with a convolutional neural network (CNN) to address the issues encountered in small-scale video quality databases with imbalanced samples and low-level feature representations for distorted videos. Transfer learning is used to enrich the distorted samples, and a CNN is used to extract high-level spatiotemporal features from the distorted videos.
4) *RecSys-DAN: Discriminative Adversarial Networks for Cross-Domain Recommender Systems:* Wang *et al*. addressed the issues of data sparsity and data imbalance in cross-domain recommender systems by leveraging the concepts from representation learning, adversarial learning, and transfer learning (particularly, domain adaptation). Unlike existing approaches, the proposed RecSys-DAN transfers the latent representations from a source domain to a target domain in an adversarial way, and it is flexible to both unimodal and multimodal scenarios, hence relatively robust to the cold-start recommendation.
5) *Siamese Neural Networks for User Identity Linkage Through Web Browsing:* Qiao *et al*. first proposed a Siamese neural network architecture-based User Identity Linkage (SAUIL) model to learn the highest level feature representation of input web-browsing behaviors for User Identity Linkage (UIL) through the Internet. Then, to address the imbalanced matching and nonmatching pairs in UIL data sets, they further proposed a cost-sensitive SAUIL (C-SAUIL) model, which assumes higher costs for misclassifying the minority class.
6) *Deep Least Squares Fisher Discriminant Analysis:* Díaz-Vico and Dorronsoro proposed deep Fisher discriminant analysis (DFDA), a straightforward nonlinear extension of least squares FDA by taking advantage of deep neural networks. For large-sample class-imbalanced data, DFDA has classification performance similar to regularized kernel FDA, but it is considerably faster in model building.

The second group of articles contains the following seven articles that introduce methods based on other machine learning paradigms for imbalanced learning:

1) *Adaptive Chunk-Based Dynamic Weighted Majority for Imbalanced Data Streams with Concept Drift:* Lu *et al*. proposed a chunk-based incremental learning method called adaptive chunk-based dynamic weighted majority (ACDWM) to deal with imbalanced streaming data

Digital Object Identifier 10.1109/TNNLS.2020.3003494

containing concept drift. ACDWM uses a dynamically weighted ensemble to address concept drift and uses statistical hypothesis tests to adaptively select the chunk size and ensure the stability of ensemble classifiers for imbalanced data streams.

2) *Discriminative Fast Hierarchical Learning for Multiclass Image Classification:* Zheng *et al*. developed a discriminative fast hierarchical learning algorithm for supporting multiclass image classification. There, a visual tree is integrated with multitask learning to achieve fast training of the tree classifier hierarchically, by handling the data imbalance and identifying the interrelated learning tasks automatically.
3) *Fast Matrix Factorization With Nonuniform Weights on Missing Data:* He *et al*. developed a fast matrix factorization algorithm to address the imbalanced learning problem to predict large-scale missing entries in a high-dimensional sparse data matrix, by weighting the missing entries nonuniformly and using elementwise alternating least squares.
4) *Iterative Privileged Learning:* Li *et al*. investigated iterative privileged learning within the context of gradient boosted decision trees (GBDTs). Unlike conventional static manipulations of privileged information, during the learning phase of the GBDT method, new decision trees are discovered to iteratively update the comments generated from the privileged information to accurately assess and coach the up-to-date model.
5) *Radial-Based Oversampling for Multiclass Imbalanced Data Classification:* Krawczyk *et al*. proposed multiclass radial-based oversampling (MC-RBO) for multiclass imbalanced learning problems. MC-RBO uses potential functions and information coming from all of the classes, for generating artificial instances, guided by the exploration of areas where the value of the mutual class distribution is very small.
6) *Self-Paced Balance Learning for Clinical Skin Disease Recognition:* Yang *et al*. proposed a self-paced balance learning (SPBL) algorithm to address not only the imbalanced sizes but also the imbalanced-recognition difficulties of multiple classes. They introduce a comprehensive metric termed the complexity of image category that is a combination of both sample number and recognition difficulty, and update the complexity by using the self-paced learning paradigm, accomplishing iterative learning of discriminative representations via balancing the complexity in each pace.
7) *Sparse Supervised Representation-Based Classifier for Uncontrolled and Imbalanced Classification:* Shu *et al*. proposed a model called Sparse Supervised Representation Classifier (SSRC) to address the fact that SRC cannot obtain satisfactory results on uncontrolled and imbalanced data sets. In SSRC, a class weight learning model is proposed to address the class-imbalance problem.

The third group includes the following two articles that present empirical comparative studies for imbalanced learning:

1) *Deep Neural Architectures for Highly Imbalanced Data in Bioinformatics:* Bugnon *et al*. provided a comparative assessment of recent deep neural architectures for dealing with the large imbalanced data issue in the classification of pre-miRNAs, along with a new graphical way for comparing classifiers' performance in the context of high-class imbalances.
2) *On the Dynamics of Classification Measures for Imbalanced and Streaming Data:* Brzezinski *et al*. studied the dynamics of eight classification measures for imbalanced and streaming data, by analyzing changes in measure values, distributions, and gradients with diverging class proportions. Their results show that the effect the class proportions have on each measure is different and should be taken into account when evaluating classifiers. They also show a direct connection between class ratio changes and certain types of concept drift.

We hope that the 15 articles selected for this Special Issue can push the research and promote many more excellent works in the field of imbalanced learning. To conclude this editorial, we would like to thank all the authors who submitted their work to this special issue, all the reviewers for their great efforts in ensuring the quality of the selected papers, and, last but not least, the Editor-in-Chief and the editorial office for their consistent support.

JING-HAO XUE, *Guest Editor*
Department of Statistical Science
University College London
London WC1E 6BT, U.K.

ZHANYU MA, *Guest Editor*
School of Artificial Intelligence
Beijing University of Posts and Telecommunications
Beijing 100876, China

MANUEL ROVERI, *Guest Editor*
Dipartimento di Elettronica e Informazione
Politecnico di Milano
20133 Milan, Italy

NATHALIE JAPKOWICZ, *Guest Editor*
Department of Computer Science
American University
Washington, DC 20016 USA

Jing-Hao Xue (Member, IEEE) received the Dr.Eng. degree in signal and information processing from Tsinghua University, Beijing, China, in 1998, and the Ph.D. degree in statistics from the University of Glasgow, Glasgow, U.U.K., in 2008.

He is currently an Associate Professor with the Department of Statistical Science, University College London, London, U.K. His research interests include statistical classification, high-dimensional data analysis, computer vision, and pattern recognition.

Zhanyu Ma (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2011.

He is currently a Full Professor with the Beijing University of Posts and Telecommunications, Beijing, China. His research interests include pattern recognition and machine learning fundamentals with a focus on applications in computer vision, multimedia signal processing, and data mining.

Manuel Roveri (Senior Member, IEEE) received the Dr.Eng. degree in computer science engineering and the Ph.D. degree in computer engineering from the Politecnico di Milano, Milan, Italy, in 2003 and 2007, respectively.

He is currently an Associate Professor with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. His current research interests include intelligent embedded and cyber-physical systems, learning in nonstationary-evolving environments, and adaptive algorithms.

Nathalie Japkowicz received the Ph.D. degree from Rutgers University, New Brunswick, NJ, USA, in 1999.

She is currently a Professor of computer science with American University, Washington, DC, USA. Her research interests are in the areas of artificial intelligence and, most specifically, machine learning, data mining, and big data analysis.